6-2021

The missing middle in product price distribution

Pao-li CHANG
Singapore Management University, plchang@smu.edu.sg

Xin YI
Singapore Management University

Haeyeon YOON

SMU ECONOMICS & STATISTICS

The Missing Middle in Product Price Distribution

Pao-Li Chang, Xin Yi, Haeyeon Yoon

Jun 2021

Paper No. 05-2021

The Missing Middle in Product Price Distribution*

Pao-Li Chang[†] Xin Yi[‡] Haeyeon Yoon[§]

June 7, 2021

Abstract

The IO literature has typically studied the supply-side factors that determine the price structure of products/services competing in a market. This paper proposes that the demand-side demographics could play an important role in shaping the product price structure. In particular, we document a "missing middle" phenomenon in both the income and the product price distributions in the U.S., based on the IPUMS ACS dataset (2005–2017) and the Nielsen Retail Scanner Data (2006–2017), for a large set of goods sold in the U.S. at the national, state, or commuting-zone level. We show that the lagged population share of the middle-income class has a positive impact on the market share (in quantity) of middle-priced varieties (and respectively so for the low/high income and price group), after controlling for product category and state (or commuting zone) fixed effects. The impacts are further stronger in commuting zones of higher population density. We then evaluate the cost-of-living implications of the observed missing-middle phenomenon, taking into account product entry, exit, and pro-competitive price effects of continuing products, in a framework that allows for non-homothetic preferences across income groups with respective to the price groups. We find that ignoring the non-homothetic demand structure and the missing-middle phenomenon understates the rise in the cost of living for the period 2006–2017. The downward bias is sizable (as large as 2 percentage points out of 11–13% increase in the cost of living for the period), and particularly noticeable for the middle-income households.

*The analyses in this paper are calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researchers and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

[†]Associate Professor, School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903. Email: plchang@smu.edu.sg. Tel: +65-68280830. Fax: +65-68280833.

[‡]Postdoctoral researcher, School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903. Email: xinyi@smu.edu.sg.

[§]Teaching Associate, School of Economics, University of Bristol, 12A Priory Rd, Bristol BS8 1TU, United Kingdom. Email: hailey.yoon@bristol.ac.uk.

Key Words: missing middle; price and income distribution; demand demographics; cost of living; entry/exit

JEL Classification: D12; D31; D61; J11; L11

1 Introduction

Sears and JC Penney, retailers whose wares are aimed squarely at middle-class Americans, are both in dire straits. Last month, Sears said it would shutter its flagship store on State Street in downtown Chicago, and JC Penney announced the closings of 33 stores and 2,000 layoffs. Loehmann's, where generations of middle-class shoppers hunted for marked-down designer labels in the Back Room, is now being liquidated after three trips to bankruptcy court since 1999. ... Investors have taken notice of the shrinking middle. Shares of Sears and JC Penney have fallen more than 50 per cent since the end of 2009, even as upper-end stores like Nordstrom and bargain basement chains like Dollar Tree and Family Dollar Stores have more than doubled in value over the same period. ... But changes in the restaurant business show that the effects of rising inequality are widespread. Foot traffic at mid-tier, casual dining properties like Red Lobster and Olive Garden has dropped in every quarter but one since 2005, according to Mr John Glass, a restaurant industry analyst at Morgan Stanley. With diners paying an average tab of US$16.50 a person at Olive Garden, he said: "The customers are middle class. They're not rich. They're not poor."

New York Times (2014b)

Among hotels, revenue per room in the high-end category, which includes brands like the Four Seasons and St Regis, grew 7.5 per cent last year, compared with a 4.1 per cent gain for mid-scale properties. ... And at General Electric, the increase in demand for high-end dishwashers and refrigerators dwarfs sales growth in mass-market models.

New York Times (2014a)

Observations such as the above made by news articles suggest that price distributions of goods/services are losing ground in the middle support, where goods/services available on the market are either very expensive catering to the super rich, or very cheap attracting bargain hunters. This phenomenon appears to be widespread and not limited to just a few sectors, as the examples above illustrate.

The IO/trade literature has typically studied the supply-side factors that determine the price structure of products/services competing in a market. This paper proposes that the demand-side factors could play an important role in shaping the product price structure. In particular, we study whether there indeed exists a widespread causal relationship between income and price distributions across time and geographical locations. The policy implications of a positive finding are potentially far-reaching, as income distribution would not only be a normative parameter to be concerned, but its extent would also affect the positive welfare outcome. In particular, it would affect firm size distributions, and products/services offered on the market and available to consumers (firms who cater to the super-rich and the poorest thrive in market shares while those who cater to the middle disappear).

As a start, we examine the price and transaction volume data for a large set of goods sold in the U.S. based on the Nielsen Retail Scanner Data (2006-2017). We document the prevalence of a "missing middle" phenomenon in the product price distribution, where the mass of the middle-priced products decreases in favor of lower-priced or higher-priced products at the barcode level. In conjunction, we document a missing-middle income phenomenon at the U.S. aggregate, state, and commuting-zone levels, based on the IPUMS ACS dataset (2005–2017). We find that the missing-middle phenomenon in product price distribution is associated with a declining population share of middle-income class. This relationship is especially pronounced in densely-populated metropolitan areas. The finding is robust to: 1) the choice of price cutoffs (tercile, quartile, or decile) that define the lower and upper bounds of the middle-price support within a product category, and 2) different resolutions of geographic units (at the state level, or at the commuting zone level) within which the price distribution for each product category is tabulated.

To assess the welfare implications of the missing-middle phenomenon, we expand on the CES demand system of Atkin, Faber and Gonzalez-Navarro (2018) (hereafter AFG), allowing for non-homothetic preferences across households of different income groups for products of different price groups. Following AFG, we decompose the cost of living into three components: the direct price-index entry effect, the exit effect, and the pro-competitive price effect. We find that the middle-priced products are most influential in determining the entry and exit effects, while the pro-competitive price effect mainly reflects the variations in the lower-priced products. During the period of analysis, new and exited varieties account for the majority of the changes in market shares. This is in contrast with the minor impacts of varieties switching (across price groups). Furthermore, the changes in market shares are largely driven by changes in the quantities (not the prices) of new/exited varieties, although products at the two tails of price support also tend to enter and exit the market frequently. Since preferences are non-homothetic, the poorer households see their cost-of-living influenced by the price dynamics of lower-priced products more (via the pro-competitive effect), while middle-income and richer households by those of middle- and higher-priced products (via the entry-and-exit effects).

Given the missing-middle phenomenon, we gauge the discrepancy in the welfare estimation between the non-homothetic CES demand system used in this study and the homothetic demand system often adopted in the literature. We find that ignoring the non-homothetic demand structure and the missing-middle phenomenon understates the rise in the cost of living (with inflation incorporated) for the period 2006–2017. The bias is sizable (up to two percentage points out of a 12% increase in the cost of living), and particularly large for middle-income households.

The remainder of the paper is structured as follows. Sections 2 and 3 introduce the measures to characterise the price and income distributions, respectively, and document the missing-middle phenomenon in both distributions. Section 4 estimates the causal effects of missing-middle in lagged income distributions on the missing-middle in price distributions across a large set of product categories, geographical units, and variations in the measure of price distributions, for the period of study. We also document heterogeneities in the strength of the causal effect by population density of the geographical units, or by other dimensions along which the mechanism could plausibly be more pronounced. Section 5 analyses the welfare impacts of the phenomena identified above. Section 6 concludes.

2 Price Distribution

This section explains how we construct market share measures for difference price groups and documents a missing middle in the price distribution over time, based on the Nielsen Retail Scanner Data (2006–2017). The Nielsen Retail Scanner data contains weekly statistics on the revenue and transaction volume for products sold in the US. This includes approximately 3.2 millions of products (identified by UPC, unique product code) in 35,000 stores under around 90 retail chains. These products are then classified into 1,075 product modules (PMC, eg., soft drinks–carbonated, or soft drinks–low calorie), 125 product groups (PGC, eg., carbonated beverages), and 10 departments (eg., dry grocery).

As a first step to tabulate the price distribution, we compute the annual average price of each UPC for a designated geographic area. This requires us to aggregate the sales revenues and volumes, respectively, of each UPC across all weeks in a year and a geographic area. The ratio of the two (= total sales / total sales volume) is then used to measure the indicative price of the UPC in a year and geographic area. We want to look at the price distribution of reasonably similar goods that are close substitutes. A *department* obviously is too aggregated a classification; for example, "dry grocery" includes "baking supplies" and "pet food", whose price and units are not comparable. Using PMCs ensures that goods are highly substitutable within a given module. However, a lot of PMCs contain fewer than 100 UPCs within the module. This renders the tabulation of some price distribution statistics susceptible to outliers. It can also be argued that some PMCs are very close substitutes with each other (e.g., "vegetables-corn-whole kernel-canned" and "vegetables-corn on the cob-canned"). Thus, we choose PGC as the unit of analysis for tabulating price distribution. The median number of UPCs within each PGC is around 1000.

The prices of all the UPCs within a PGC are then used to tabulate the price distribution of a PGC, using the sales volumes (physical units such as ounce) of each UPC as weights.[1] A PGC is dropped from the analysis if it contains fewer than 100 effective

[1] In particular, we find the mode of the physical unit of measure for each PGC and year, and keep only UPCs whose physical unit of measure is the same as the mode. This implies that our analysis is based on the most representative types of products that are similar enough to have the same type of quantity measure within each PGC. We also drop all the PMCs (and hence all UPCs under them) that are marked as "Deferred" and whose information entry may be outdated according to the Nielsen data manuals.

observations of UPCs. Following this procedure, we obtain a panel of price distribution statistics for the period 2006–2017 for 119 PGCs for each definition of geographic area.

2.1 Measures

In particular, we characterise the price distribution for each PGC in a year-area by the market shares (in terms of sales volume in physical units) of UPCs that fall within a certain price range, where the price cutoffs are fixed in terms of the 2006 price distribution of the same PGC in the same area. For example, in our baseline analysis, we define the cheap products as those UPCs whose deflated real price is below the 33rd percentile of the year-2006 price distribution (of the corresponding PGC g in area s).[2] Correspondingly, we define the expensive products as those UPCs with deflated real price above the 66th percentile of the 2006 price distribution. The middle-priced products are defined as those with real prices in between the two terciles. Given this, the market share of UPCs that fall in a particular price range, $Share_{st,g,p}$, is an area (s), year (t), PGC (g), and price-range (p) specific variable, such that:

$$Share_{s,t,g,p} = \sum_{u \in \Omega_g(p)} \left(Share_{s,t,u} \right), \tag{1}$$

where $Share_{s,t,u} = Sales_{s,t,u}/Sales_{s,t,g}$ is the market share of UPC u in a PGC g in an area and year (s,t); and $\Omega_g(p)$ is the set of UPCs whose price falls within the price range $Price_{s,g,p}$ characterized by price range p. Note that by definition, the market shares are 33% for low-, middle-, and high-priced products for any PGC-area pair in the base year (i.e., 2006). In robustness checks, we also use alternative thresholds such as 25th/75th and 10th/90th percentiles to define the lower/upper price cutoffs for the middle-priced products.

2.2 Missing Middle in Price Distribution

We now document the missing middle phenomenon in the price distribution, by examining the time trend of the market shares of middle-priced products in the U.S. and in its sub regions. In particular, we scatter-plot the changes in market shares $Share_{s,t,g,p}$ of each price group ($p = \{L, M, H\}$) for each PGC (g) relative to the base year 2006, according to alternative lower/upper price cutoffs (T for terciles, Q for quartiles, and D for deciles) for the middle-priced products. The results are shown in Figure 1 for all sales across the U.S..

[Insert Figure 1 Here]

[2] We deflate the nominal price to 1999 real dollars by using the CPI data from the IPUMS database.

Figure 1 indicates that middle range of price distributions gradually lost its support during the period 2006–2017. The lost market shares for the middle-priced products are absorbed by the increase in both low- and high-priced products. The pattern is more pronounced when we use a more general definition (e.g., the middle 50% and the middle 80%) for the middle-price support. This suggests that the majority of the decline in the market share of the middle-priced products happened at the two tails of the middle-price support, in that varieties that are relatively cheap and expensive dropped out from the middle-price support.

Figures 2–4 tabulate the price distribution for each PGC in 2017 (relative to 2006) at the state level (for each of the 48 contiguous states and the District of Columbia, for which the price data are available). The pattern suggests that the missing middle phenomenon is prevalent across states, and is more pronounced as we use a broader definition of middle support (from terciles, to quartiles, to deciles).

[Insert Figures 2–4 Here]

In Figure 5, we further narrow down the geographical area, and calculate the market share of L/M/H-priced products for each PGC at the level of commuting zones. Given the large number of commuting zones, we summarize the results by the average of the market shares across commuting zones within the same bin of population density in 2017 (relative to 2006).[3] In particular, commuting zones are ranked by their population densities, weighted by the population size of the commuting zone, and grouped into 10 bins (with each bin representing approximately 10% of the population). Bin 10 (1) has the highest (lowest) population density. The figure similarly indicates a decrease in the market share of the middle-priced products. The pattern appears to be more pronounced in commuting zones of higher population density, and as observed above, when we use a broader definition of middle-price support.

[Insert Figure 5 Here]

3 Income Distribution

3.1 Measures

We use the IPUMS ACS dataset (2005–2017) to construct statistics of the income distribution for the U.S. and its geographical units (states/district, and commuting zones) in each year. The sampling unit is the household and all persons residing in the household. We keep observations with positive, non-missing household income data, and drop

[3] We compute the population density for each CZ, using a crosswalk from the county level to the CZ level provided by Autor and Dorn (2013), the county population in 2005 provided by the NBER, and the county land areas provided by the Census Bureau.

'households' that are group quarters (such as institutions and boarding houses). Since the Nielsen Retail Scanner data characterized in Section 2 cover only the contiguous states of the U.S., we drop the ACS income observations for Alaska and Hawaii correspondingly.[4] The nominal income is deflated to 1999 real dollars by using the CPI data from the IPUMS database (as for the nominal price in Section 2).

We consider two alternative approaches of translating household income to individual income. In the benchmark reported below, we simply divide the gross income of the household by the number of household members:

$$\text{Per-capita income} = \frac{\text{Gross household income}}{\text{Number of persons}}. \tag{2}$$

As a robustness check, we also consider adjusted per-capita income, following Kochhar and Fry (2015) of Pew Research Center. The findings based on this alternative approach are similar to the benchmark, and are reported in Appendix B.

Table 1 reports the summary statistics of the IPUMS ACS sample for the period 2005–2017. For each household and its household income per capita, it is attached a weight (= number of persons residing in the same household) reflecting how many individuals in the U.S. population that are represented by this sampled household income per capita. The percentile measures of the income distribution are computed incorporating these weights.

[Insert Table 1 Here]

Given the individual income observations constructed above, we measure the share of middle-income population, according to the definition used by Kochhar and Fry (2015) of Pew Research Center. In particular, we define the middle-income class as individuals whose income falls in between 2/3 and twice of the median income level.[5] Correspondingly, the population share of the lower-income class is defined as the share of population with an income below 2/3 of the median income level, while that of upper-income class the share of population whose income exceeds twice the median income level.

3.2 Missing Middle in Income Distribution

We now report the time-series of the population share of different income classes at the U.S. national level. Figure 6 suggests that the middle-income population share has been declining since 2008 and remains low afterwards. The decline in the middle-income population share could be as large as 2 percentage points on a national scale (and 3

[4] We also drop household observations with missing census tract identifiers, because they are required to identify the commuting zone of the household (for tabulating income distribution at the level of commuting zones).

[5] Tayor et al. (2008) of Pew Research Center has used an alternative definition, based on 75% and 150% of the median income as the cutoffs, while Krueger (2012) used 50% and 150%, respectively. The findings of the paper are robust to these alternative definitions.

percentage points if based on adjusted per-capita income as illustrated in Figure A.1). In the meantime, both the lower- and upper-income population shares have been rising, although the upper-income share rises slightly more than the lower-income share (again, the variations are larger if based on adjusted per-capita income). The pattern echoes that of the decline in the market share of middle-priced products as documented in Section 2.

[Insert Figure 6 Here]

Next, we examine the income distribution at the state level. We use the national median income in classifying income class, and construct the share of populations belonging to different income classes at the state level. Figure 7 reports the time-series of the population share of the L/M/H-income class for each state/district. While the majority of the states have seen a decline in the middle-income population share, some states experienced different patterns in the L/H-income shares in contrast to the national trend. In some states/district such as the District of Columbia and Pennsylvania, both the middle- and lower-income population shares declined. The opposite is the case for states such as Michigan, whose population share of the lower-income group has increased while those of M/H-income groups have dwindled. It is worth noting that the missing-middle phenomenon in income distribution is more pronounced in states that are in close proximity to major metropolitan areas, such as Connecticut, Massachusetts, Maryland, and New Jersey.

[Insert Figure 7 Here]

We examine below whether the pattern of population shares of income groups varies with urban density systematically over the years. As documented by Autor (2019), the population share of middle-skilled jobs in the U.S. has been decreasing in the last four decades, and the phenomenon was particularly evident for the urban area. Given that income is positively correlated with skills, we verify whether the missing-middle in income distribution is also more conspicuous in metropolitan areas.

We first convert the income data of ACS with census tract identifiers at the level of Public Use Microdata Areas (PUMAs) to the level of commuting zones (CZ), which is widely considered as good proxies for local labor markets. The problem at hand is that a PUMA sometimes overlaps with multiple commuting zones. Autor and Dorn (2013) propose a solution to this issue by computing the probability that an individual living in a PUMA is from a particular CZ. We utilize these probabilities,[6] and assign the individual income observations (household income per capita, weighted by household size) from PUMAs to CZs. Table 2 provides the summary statistics of income distribution in 2005 for commuting zones sorted into 10 bins of population density.

[Insert Table 2 Here]

[6]These are available from David Dorn's website.

Table 2 indicates that income distributions are heterogeneous across commuting zones of different population densities. First, Columns 2–3 suggest that the average income is higher in more densely populated metropolitan areas. This is not surprising, considering the vast literature on how urbanization facilitates economies of scale and knowledge spillover. Correspondingly, Columns 4–6 suggest that commuting zones of higher population density are associated with a larger share of upper-income population and a smaller share of lower-income population.

We now document the pattern of income distribution across time by the population density of commuting zones. Figure 8 suggests that the declining population share of the middle-income group is more severe in more densely populated commuting zones. Second, the missing middle-income class over time is associated with the rise in *both* lower- and upper-income population shares; and the pattern is more pronounced in the more densely populated areas. Our findings above are consistent with and reinforce the message of Autor (2019) that the middle-skill employment share in the U.S. has declined, and more so in the more densely populated commuting zones, over the past decades.

[Insert Figure 8 Here]

4 Estimations

In Sections 2–3, we have documented a shrinking middle in both the price and income distributions, in the sense that the market share of middle-priced products and the population share of the middle-income class have declined during the period 2006–2017 (2005–2017). We now attempt to connect the two phenomena by arguing that the changes in product price distribution are driven by the changes in income distribution. The main idea is to regress the market shares of different price groups on the lagged population shares of different income groups, controlling for a battery of fixed effects (except for those that would absorb the time variation, the main identifying variation that we will exploit). In essence, our baseline estimation specification is as follows:

$$Share_{s,t,g,p} = \alpha_1 Frac_{s,t-1,y} + \gamma_g + \delta_s + \varepsilon_{s,t,g,p} \tag{3}$$

where $Share_{s,t,g,p}$ is the market share of price group p, of PGC g, in area s at time t; $Frac_{s,t-1,y}$ is the population share of income group y in area s at time $t-1$, where $p, y \in \{L, M, H\}$; γ_g and δ_s are PGC- and area-specific fixed effects; and $\varepsilon_{s,t,g,p}$ is the residual error term. In the benchmark, we use the terciles (the 33rd/66th percentiles of the base-year 2006 price distribution) as the cutoffs for defining the middle-priced products in each PGC in each year $t \in \{2007, \ldots, 2017\}$. We also provide robustness checks based on quartiles and deciles. More details on the market share measure $Share_{s,t,g,p}$ are provided in Section 2.1.

4.1 State-Level Results

We first report the estimation results when the market share of L/M/H-priced products and the population shares of L/M/H-income class are defined at the state level. The benchmark result based on terciles for price cutoffs, and the robustness checks based on quartiles/deciles, are reported in Table 3, and Tables 4/5, respectively.

[Insert Tables 3–5 Here]

The estimates of the slope coefficient α_1 are positive and significant (economically and statistically) in all specifications. This suggests that the demand-side demographics have had important impacts on the product market shares. In particular, Table 3 suggests that on average across PGCs, states and years, a 10 percentage point decrease in the population share of the middle-income class leads to a 3.62 percentage point decrease in the market share of the middle-priced products, after controlling for the state- and PGC-specific fixed effects. In addition, Table 3 indicates that the effect of demographics on product demand is indeed the strongest for the higher-income class, and the weakest for the lower-income class. This could imply that the income elasticity is larger for higher-priced products. It also echoes the findings in Jaravel (2019) that there tended to be more net entry of higher-priced products, and Argente and Lee (2021) that higher-income households tended to experience smaller inflation in their consumption baskets. Our analysis, however, provides further insights on the overall changes in the market place of household consumer goods, when combined with the missing-middle phenomenon documented in Sections 2–3. In particular, our analysis suggests that the market place of middle-priced products has undergone changes in a direction opposite of the higher-/lower-priced products and shrunken in its mass in response to the shrinking consumer base for such products.

Table 4, based on the 25th/75th percentiles for price cutoffs, reports a similar qualitative pattern as observed above. The effect estimates are however stronger for the M/H-priced products in response to the M/H-income population shares, and weaker for the L-priced products. Table 5, based on the 10th/90th percentiles for price cutoffs, indicates that the effects of population share on market share further weaken for L-priced products, but now also for the M/H-priced products. This could imply that the income cutoffs (2/3 and twice of the median income level) we use to classify the middle-income class match the tercile (or quartile) price cutoffs better in terms of the income-consumption nexus, such that a further widening in the defined price support (to the range of 10%–90%) for the middle-priced products weakens the relationship between the consumer base and the product market share identified in the benchmark.

4.2 Commuting-Zone-Level Results

We further disaggregate the geographical area and examine the specification in (3) at the commuting-zone level. We verify whether the hypothesis (that changes in the income distribution affect the price distributions) still holds in the context of commuting zones. The results are reported in Tables 6, 7, and 8 for the tercile, quartile, and decile price cutoffs, respectively.

[Insert Tables 6–8 Here]

The results at the commuting-zone level are largely consistent with those of state-level regressions. All coefficient estimates of α_1 are positive and significant, although the average effects of the M-income population share on the M-priced product share tend to be smaller at the commuting-zone level. We explore further this issue in the following section. Finally, the finding at the state level that the effect estimates of α_1 attenuate when we further broaden the definition of the middle-priced products from 25th/75th to 10th/90th percentiles, remains to hold overall at the finer commuting-zone level, suggesting a poorer match between income class and consumption baskets as we adopt too broad a middle-price support.

4.3 Heterogeneity Effects by Population Density

We now examine whether the effect of income distribution on price distribution varies systematically with population density of the commuting zones. The baseline specification is amended to incorporate an interactive term between the population density of CZ s in 2005 and the population shares of an income group in CZ s in year t such that:

$$Share_{s,t,g,p} = \alpha_1 Frac_{s,t-1,y} + \alpha_2 \left(Frac_{s,t-1,y} \times \mathbf{1}\{Density_s \geq median\}\right) + \gamma_g + \delta_s + \varepsilon_{s,t,g,p}, \tag{4}$$

where $Density_s$ is the population density of CZ s in 2005, and the indicator $\mathbf{1}\{Density_s \geq median\}$ equals one if the population density of CZ s is above or equal to the national median (across all 722 commuting zones) in 2005. Thus, the coefficient α_1 measures the marginal effect of income population shares on product market shares for the commuting zones with below-the-median population density, and $(\alpha_1 + \alpha_2)$ the corresponding effect for the commuting zones with above-the-median population density. A positive (negative) α_2 implies that an income group's population share has stronger (weaker) effects in more densely populated metropolitan areas. An insignificant estimate of α_2 implies that the effects are uniform regardless of the commuting zone's population density.

[Insert Tables 9–11 Here]

The results are reported in Tables 9, 10,and 11 for the tercile, quartile, and decile price cutoffs, respectively. First, the estimates of α_1 and α_2 are both positive and significant for the middle-/higher-income group. This implies that the declining (increasing) population share of middle (upper) income class reduces (increases) the market shares of middle-priced (higher-priced) products; and these effects are more pronounced in more densely populated areas. This finding echoes the recent work in the agglomeration literature where the agglomeration benefits of larger cities are known to be biased towards higher-skill workers. It could be that middle-skilled workers or middle-income workers in larger cities face larger setbacks in their wages at the intensive margin (in addition to a decrease in the number of such workers at the extensive margin), such that the reduction in the demand for middle-priced products is more pronounced in larger cities. Our findings add novel insights to the literature, in the sense that we are the first study to highlight the spatial (agglomeration) dimension in how the demand-side demographics influences product price distributions. Since the level effects (α_1) of M-income population shares on M-price product shares are relatively small, in comparison with the interactive effects (α_2), this suggests that the impact of the missing-middle phenomenon is largely a large-city concern. On the other hand, the level effects for the H-price (H-income) group are just as economically substantial as its interactive effects. Thus, the changes in the share of the H-priced products (due to demographic changes in income shares) are relatively prevalent across commuting zones of all population densities.

Next, the net effect ($\alpha_1 + \alpha_2$) of L-income population shares on L-priced product market shares is also positive on average across all commuting zones. However, the net effect is smaller in more densely populated commuting zones by the tercile price cutoffs, the same across commuting zones by the quartile price cutoffs, and larger in more densely populated commuting zones by the decile price cutoffs. The mechanism is not entirely clear to us. Our conjecture is that although the population shares of L-income workers increases in general over the years (at the extensive margin), their wages might not have improved as much (at the intensive margin) in the more densely populated areas, such that their aggregate purchasing powers are weaker relatively in the more densely populated areas, which was manifested in an increased demand for extremely cheap products (under the 10th percentile) but not for products priced in between the 10th and 33rd percentiles.

5 Welfare Implications

We now examine the welfare implications of the missing-middle phenomenon for the U.S. consumers. To this end, we modify the CES demand system of Atkin, Faber and Gonzalez-Navarro (2018) to allow for non-homothetic preference for L/M/H-priced products within each PGC. One advantage of adopting this approach is that it enables us to quantify and decompose the welfare effects (in terms of the cost of living) into three components: the direct price-index "entry" effect, the exit effect, and the pro-competitive price effect, with each component further disaggregated by the contributions due to products of the L/M/H-price group, respectively. This allows us to examine how the missing middle in price distributions influences the welfare of the U.S. consumers.

5.1 Setup

Specifically, we assume a two-tier demand system. In the top tier, we assume that preference is non-homothetic such that for an individual in income group y,

$$U_y = \prod_g \prod_p [Q_{ygp}]^{\theta_{ygp}} \tag{5}$$

where Q_{ygp} is the consumption of products in price group p of product category g by individuals in income group y, and θ_{ygp} is the corresponding Cobb-Douglas expenditure share. Unlike the conventional Cobb-Douglas demand system over different product categories (such as those in Atkin, Faber and Gonzalez-Navarro, 2018), we assume that the expenditure share is price-group *and* product-category specific. This introduces non-homothetic demand into consumer preferences in a transparent and data-driven manner. Preferences are non-homothetic if the expenditure shares differ systematically across income groups. We verify this using the Nielsen Consumer Panel data (2005–2017), which record the consumption behavior of households (the barcode-level products purchased) and the characteristics of the household (including household income and household size). More details about the dataset are provided in Section 5.2.

[Insert Figure 9 Here]

Figure 9 plots the times series of the expenditure shares by each income group on products in the L/M/H-price group (defined at the PGC level), pooling observations across PGCs and the U.S. (48 contiguous states and the District of Columbia). The pattern suggests that the expenditure shares differ systematically across income groups with respect to price groups. The expenditure share on lower-priced products is the highest by the lower-income group and the lowest by the higher-income group. The pattern reverses for the higher-priced products: the expenditure share on these expensive products is the highest by the rich individuals and the lowest by the poor individuals. These differences in the spending pattern are economically significant in the sense that the difference in the expenditure shares allocated to a given price group of products by the poor and the rich could be as large as 10 percentage points. The difference in expenditure shares on middle-priced products across income groups are, in comparison, not as pronounced, although it is shown that the higher-income group tends to spend less (in terms of expenditure shares) on middle-priced products than the other two income groups. Taken together, Figure 9 provides empirical support for the non-homothetic demand structure assumed in (5). This assumption will also prove to be instrumental in our analysis below.

Next, the bottom tier of the consumer preference is assumed to be CES over different products (varieties) defined at the bar-code (UPC) level such that:

$$Q_{ygp} = \left[\sum_{u} \beta_{ygp,u} \left(q_{ygp,u} \right)^{\frac{\sigma_{ygp}-1}{\sigma_{ygp}}} \right]^{\frac{\sigma_{ygp}}{\sigma_{ygp}-1}}, \tag{6}$$

where $\beta_{ygp,u}$ is a taste parameter attached by income group y to product (variety) u in price range p of PGC g, $q_{ygp,u}$ the quantity consumed by income group y of variety u in price range p of PGC g, and $\sigma_{ygp} > 1$ is the elasticity of substitution across varieties in price range p of PGC g specific to income group y.

Following similar derivations as in Atkin, Faber and Gonzalez-Navarro (2018), we can show that the change in the cost of living in any year t (relative to the base year 0) is given by:

$$\prod_{g} \prod_{p} \left(\left(\frac{\sum_{u \in u^{c,t}_{ygp}} \phi^{t}_{ygp,u}}{\sum_{u \in u^{c,t}_{ygp}} \phi^{0}_{ygp,u}} \right)^{\frac{1}{\sigma_{ygp}-1}} \prod_{u \in u^{c,t}_{ygp}} \left(\frac{\pi^{t}_{gp,u}}{\pi^{0}_{gp,u}} \right)^{\omega^{t}_{ygp,u}} \right)^{\theta_{ygp}} - 1, \tag{7}$$

where $\omega^{t}_{ygp,u}$ is the ideal log change weight,[7] $\phi^{t}_{ygp,u}$ is the market share (in terms of sales revenues) of variety u in the price-group p of product-category g in year t, $\pi^{t}_{gp,u}$ is the price of the corresponding variety u in year t, and $u^{c,t}_{ygp}$ is the set of continuing products (relative to the base year 0) in the price-group p of product-category g in year t consumed by income group y. Readers who are familiar with the literature would see that, three components have important influence on the magnitude of changes in the cost of living. First, the expression $\prod_{u \in u^{c,t}_{ygp}} \left(\pi^{t}_{gp,u} / \pi^{0}_{gp,u} \right)^{\omega^{t}_{ygp,u}}$ captures how prices of continuing products change over time. This could be due to either changes in the competitive environment or general inflationary pressures. Second, the expression $\sum_{u \in u^{c,t}_{ygp}} \phi^{t}_{ygp,u}$ in the numerator of the first term captures the effect of entry of new products. Notice that the sum of market shares of continuing products is exactly one minus the market share of new products from the perspective of year t. If this sum decreases, it implies more entries by new products and a positive impact on welfare (via a reduced cost of living). Third, the expression

[7] Specifically,

$$\omega^{t}_{ygp,u} = \left(\frac{\tilde{\phi}^{t}_{ygp,u} - \tilde{\phi}^{0}_{ygp,u}}{\ln \tilde{\phi}^{t}_{ygp,u} - \ln \tilde{\phi}^{0}_{ygp,u}} \right) \Bigg/ \sum_{u \in u^{c,t}_{ygp}} \left(\frac{\tilde{\phi}^{t}_{ygp,u} - \tilde{\phi}^{0}_{ygp,u}}{\ln \tilde{\phi}^{t}_{ygp,u} - \ln \tilde{\phi}^{0}_{ygp,u}} \right), \tag{8}$$

where $\tilde{\phi}^{t}_{ygp,u} = \pi^{t}_{gp,u} q^{t}_{ygp,u} \Big/ \sum_{u \in u^{c,t}_{ygp}} \pi^{t}_{gp,u} q^{t}_{ygp,u}$ measures the market share (in terms of sales revenues) of variety u in the set $u^{c,t}_{ygp}$.

$\sum_{u \in u_{ygp}^{c,t}} \phi_{ygp,u}^0$ in the denominator of the first term indicates how the exit of products affects the consumer welfare. This expression corresponds to one minus the market share of exited products from the perspective of base year 0. A smaller number of the term implies more product exits and a negative impact on welfare (via an increased cost of living). To isolate the effect of each of these components on the cost of living, we follow Atkin, Faber and Gonzalez-Navarro (2018) and decompose (7) as follows:

$$\prod_g \prod_p \left(\left(\frac{\sum_{u \in u_{ygp}^{c,t}} \phi_{ygp,u}^t}{\sum_{u \in u_{ygp}^{c,t}} \phi_{ygp,u}^0} \right)^{\frac{1}{\sigma_{ygp}-1}} \prod_{u \in u_{ygp}^{c,t}} \left(\frac{\pi_{gp,u}^t}{\pi_{gp,u}^0} \right)^{\omega_{ygp,u}^t} \right)^{\theta_{ygp}} - 1$$
$$= \text{Direct Price-Index “Entry” Effect} + \text{Exit Effect} + \text{Pro-competitive Price Effect},$$

where the individual components are, respectively:

(I) Direct Price-Index “Entry” Effect

$$= \prod_g \prod_p \left(\left(\frac{\sum_{u \in u_{ygp}^{c,t}} \phi_{ygp,u}^t}{\sum_{u \in u_{ygp}^{c,t}} \phi_{ygp,u}^0} \right)^{\frac{1}{\sigma_{ygp}-1}} \prod_{u \in u_{ygp}^{c,t}} \left(\frac{\pi_{gp,u}^t}{\pi_{gp,u}^0} \right)^{\omega_{ygp,u}^t} \right)^{\theta_{ygp}}$$
$$- \prod_g \prod_p \left(\left(\frac{1}{\sum_{u \in u_{ygp}^{c,t}} \phi_{ygp,u}^0} \right)^{\frac{1}{\sigma_{ygp}-1}} \prod_{u \in u_{ygp}^{c,t}} \left(\frac{\pi_{gp,u}^t}{\pi_{gp,u}^0} \right)^{\omega_{ygp,u}^t} \right)^{\theta_{ygp}}, \tag{9}$$

(II) Exit Effect

$$= \prod_g \prod_p \left(\left(\frac{1}{\sum_{u \in u_{ygp}^{c,t}} \phi_{ygp,u}^0} \right)^{\frac{1}{\sigma_{ygp}-1}} \prod_{u \in u_{ygp}^{c,t}} \left(\frac{\pi_{gp,u}^t}{\pi_{gp,u}^0} \right)^{\omega_{ygp,u}^t} \right)^{\theta_{ygp}}$$
$$- \prod_g \prod_p \left(\prod_{u \in u_{ygp}^{c,t}} \left(\frac{\pi_{gp,u}^t}{\pi_{gp,u}^0} \right)^{\omega_{ygp,u}^t} \right)^{\theta_{ygp}}, \tag{10}$$

(III) Pro-competitive Price Effect

$$= \prod_g \prod_p \left(\prod_{u \in u_{ygp}^{c,t}} \left(\frac{\pi_{gp,u}^t}{\pi_{gp,u}^0} \right)^{\omega_{ygp,u}^t} \right)^{\theta_{ygp}} - 1. \tag{11}$$

The interpretations of these components are as follows. First, the direct price-index “entry” effect is a non-positive term since $\sum_{u \in u_{ygp}^{c,t}} \phi_{ygp,u}^t \leq 1$ holds. As discussed above, the sum $\sum_{u \in u_{ygp}^{c,t}} \phi_{ygp,u}^t$ is further smaller than 1, when the share of expenditure consumers spend on new products is larger. Entry of new products reduces the cost of living for the consumers, given the love-of-variety assumption embedded in the CES preference structure. Thus, the cost of living falls by more (the effect of entry on the cost of living is more negative), when there are more product entries. Second, the exit effect is nonnegative, because the expression $\left(1/\sum_{u\in u^{c,t}_{ygp}} \phi^0_{ygp,u}\right) \geq 1$ holds. As more products were dropped from the marketplace, the expression becomes larger and increases the cost of living. Intuitively, with more product exits, consumers have fewer varieties to choose from, and by the love-of-variety preference structure, increase the price index and the cost of living. Third, the interpretation of the price effect is straightforward, as it simply tracks how the prices of continuing varieties evolve over time. Movement in variety prices affects the cost of living in the same direction, weighted by their ideal log change weights ($\omega^t_{ygp,u}$) within the set $u^{c,t}_{ygp}$ and the Cobb-Douglas expenditure shares (θ_{ygp}).

5.2 Data and Measurements

We compute the changes in the cost of living, and its components, using the Nielsen Consumer Panel data for the period 2005–2017. This dataset records information on the household demographics, the trips to retail stores, the purchases at the UPC (bar-code) level in each trip, and the product characteristics (the product's PMC, PGC, physical unit of measure, etc.). Summary statistics for this dataset are provided in Table 12. This dataset is useful for the current set of analysis, because it provides direct observations on the links between household demographics and consumption patterns. Note, however, that the sample of households is much smaller than the IPUMS ACS (about 6% of the latter), and the household income reported are banded (rather than in exact amounts). Similarly, the set of products sampled (reported by the households sampled) are also much smaller than the Nielsen Retail Scanner Data. This explains our choice of the Nielsen Retail Scanner Data (2006–2017) and the IPUMS ACS data (2005–2017) when we characterised the product price and income profiles in Sections 2-4.

In particular, we construct the price groups for each product category (PGC) following a similar methodology as in Section 2 for the Nielsen Retail Scanner Data. The missing middle pattern in price distributions is robust to the change of dataset. Next, as noted above, the Nielsen Consumer Panel data only report the band of household income (e.g., \$30,000 to \$50,000). We proxy the household income by the lower bound of the income band that the household belongs to. As such, the actual household income is no smaller than the proxy income. We use the household income and household size information to compute a proxy of per-capita household income. We then use this per-capita income measure to assign households to income groups, based on the cutoffs that we previously constructed from the IPUMS ACS data. Finally, we compute the changes in the cost of living and its components, using: (i) the imputed price for each variety u in the price-group p of product-category g in year t, (ii) the market share and the ideal log change weight of the corresponding variety, specific to income group y, and (iii) the Cobb-Douglas shares for each combination of ygp.[8] We calibrate the elasticity of substitution to 5, which

[8]The Cobb-Douglas shares are time-invariant in the model; we take the average across years of the

is in line with the literature estimates that use the Nielsen datasets.

5.3 Cost of Living

The changes in the cost of living for the period 2005–2017, according to (7), and its decompositions, according to (9)–(11), are summarized in Figure 10. For the benchmark, we do not deflate prices, so that changes in the cost of living also reflect the inflation over years. The results based on deflated prices (in 1999 real dollars) are reported in the appendix. Figure 10 indicates that changes in the cost of living over time in the U.S. is biased in favor of the richer households, consistent with the previous literature (Argente and Lee, 2021; Jaravel, 2019). While the changes in the cost of living predominantly reflect the price effect (when inflation is taken into account), there is also a net entry effect over time across all households. In particular, the net entry effect is also stronger for the richer households, which again is consistent with the literature, in that there have been more product innovations (net of product exits) tailored towards the taste of richer households (Jaravel, 2019).

[Insert Figure 10 Here]

5.4 Margins of Sales: Number of Varieties, Quantities, and Prices

In Section 2, we document the pattern of missing middle in price distributions, in terms of quantities sold by each price group. In calculating the changes in the cost of living, the market shares in terms of sales value are the basis for measuring the entry/exit effects (and the ideal log change weight for the pro-competitive price effect). We now characterize the set of varieties that enter and exit a price group, and examine the relative importance of the quantity versus price mechanisms in determining the market shares.

Given our set up, because entry and exit in our demand system is price-group specific (within yg), it is entirely possible that a product/variety does not disappear from the consumption basket of yg, but is dropped from a price group when its price changes so much that it "switches" to another group. In this section, to facilitate discussions, we reserve the terms "entry" and "exit" for products/varieties that enter or disappear from the consumption basket of yg, while using the term "switching products" for those that switch across price groups but remain within the same consumption basket of yg.

We construct our statistics as follows. First, we compute the proportion of entry/exit/switching varieties (defined at the yg level) for each price-group and income-group pair. For example, a H-priced variety in the consumption basket of yg in year t

expenditure shares in the data for each combination of ygp.

could be a new variety (not present in the consumption basket of yg in 2005) or a switching variety that were in the consumption basket of yg in 2005, but switched from the M-price group to the H-price group. Conversely, a M-priced variety in the consumption basket of yg in year 2005 could disappear from the consumption basket of yg in year t. We construct the proportion of entry/switch-in varieties using the number of these varieties divided by the total number of varieties in the current year t for each price-income group, by pulling across all PGCs. Similarly, we construct the proportion of exit/switch-out varieties using the number of these varieties divided by the total number of varieties in the base year 2005 for each price-income group, by pulling across all PGCs. Second, we compute the quantity share of entry/switch-in (exit/switch-out) products for each price-income group and PGC, relative to the current year t (the base year 2005), and take the average of the shares across all PGCs. This is in view that the physical units of measure are not the same across PGCs (while all varieties within a PGC have the same physical unit of measure by construction, cf. Section 2). Third, we then calculate the expenditure share of entry/switch-in (exit/switch-out) products for each price-income group, by pulling the expenditures across PGCs. These results are summarized in Figures 11–16.

[Insert Figures 11–16 Here]

Figures 13 and 16 indicate that product entries/exits account for increasing larger expenditure shares over time, relative to product switching. For example, for middle-income households, of their M-priced consumption basket in 2017, the new (switched-in) products relative to 2005 account for close to 67% (13%) of expenditures, with the remaining 20% accounted for by varieties that remain in their consumption basket since 2005. On the other hand, of their M-priced consumption basket in 2005, the exit (switched-out) products account for close to 61% (16%) of expenditures.

Next, the pattern of expenditure shares follows that of quantity shares closely (especially for the M-price group). However, the gap in expenditure shares across price-groups tends to be larger than the gap in quantity shares. For example, the expenditure shares of entry/switch-in varieties of the H-price group (cf. Figure 13) are on a higher scale than the corresponding quantity shares (cf. Figure 12), while the expenditure and quantity shares of entry/switch-in varieties of the M-price group are nearly the same. As a result, the gap between the H-price and M-price groups is wider in terms of expenditure than quantity. A similar observation can be made for product exit/switching-out (cf. Figure 16 versus Figure 15). This suggests that among the H-priced products, it is the most pricey varieties that are entering/leaving the market. Similarly, among the L-priced products, it is the cheapest varieties that are entering/leaving the market.

Further, the varieties that enter/exit the markets are those with relatively small quantities. We arrive at this conclusion by comparing the proportion of these varieties with the quantity shares of the corresponding varieties (Figure 11 versus Figure 12, and Figure 14 versus Figure 15), where the proportion of varieties that enter/exit the markets is far larger than the quantity share of the corresponding varieties. This feature is consistent with the Melitz (2003)-type model in that firms around the entry/exit productivity cutoff are the least efficient ones (and produce the least in terms of quantities). Nonetheless, features (such as product qualities) beyond the conventional Melitz-type model are required to also accommodate the pattern found above that products/varieties that leave the market in the L-price group also tend to be the lowest-priced products.

5.5 Bias in the Measurement of Cost of Living

To fully quantify how the welfare implications are affected by the presence of non-homothetic demand with respect to price groups, the missing-middle phenomenon in price distributions, and the presence of product-switching across price groups, we now compare the results based on the demand structure proposed in Section 5.1 with a restricted version that assumes away non-homothetic preferences with respect to price groups. In particular, we change the upper tier of the demand system in Section 5.1 to the following conventional setup:

$$U_y = \prod_g [Q_{yg}]^{\theta_{yg}}, \tag{12}$$

$$Q_{yg} = \left[\sum_u \beta_{yg,u} \left(q_{yg,u}\right)^{\frac{\sigma_{yg}-1}{\sigma_{yg}}} \right]^{\frac{\sigma_{yg}}{\sigma_{yg}-1}}, \tag{13}$$

where the Cobb-Douglas expenditure shares are not different across income groups with respect to price groups, and a variety u is not attached with a price-group label. We compute the cost of living and its decomposition under this alternative preference structure, and compare them with those derived in Section 5.3. The results are summarized in Figures 17–21.

[Insert Figures Here]

In general, the conventional demand system understates the increase in the cost of living, especially for middle-income households during 2012–2017. This downward bias is quantitatively significant and could be as large as 2 percentage points (out of 11–13% increase in the cost of living for the period). The bias reflects to a large extent the biases in the entry and exit effects. Intuitively, our demand system with non-homothetic preferences (with respect to price groups) implies greater entry and exit effects (in absolute terms), because of product-switching across price groups. Nonetheless, the net entry effect is overstated in the conventional setup, and more so for middle-income (and lower-income) households. This reflects the bias in measuring changes in the cost of living by omitting the missing-middle phenomenon and its stronger impacts on middle-income (and lower-income) households in the presence of non-homothetic preferences across price groups (cf. Figure 9).

6 Conclusion

The research in this paper has several potential policy implications and extensions. First, given any initial shocks to the income distribution, the paper's findings imply amplified impacts on welfare — due to the direct income effect, but also the indirect impact on the cost of living as a result of changes in the set of products available to the different income groups. For example, our research provide new insights into the potential ramifications of the ongoing Covid-19 global pandemic. As we document in the paper, the population share of the middle-income class have decreased over the recent decade, and this pattern is especially pronounced during and in the aftermath of the Great Recession of 2008–2011. We are witnessing a similar episode as the world is hit by the Covid-19 global pandemic. Bloomberg (2021b) reported that millions of the global middle class, especially those from the developing world, are being forced into poverty by the Covid pandemic. Similarly, Bloomberg (2021a) reported that in the U.S., the road to recovery is hugely uneven, depending on the access to financial markets. The richer households have become richer thanks to the soaring asset prices backed by the ultra-loose Fed monetary policies, while the poorer households have fallen behind. Our research suggests that the current hollowing out of the middle-income class could lead to systematic changes in the price distribution of consumer goods, which could further amplify the biased welfare implications against the middle- and lower-income class, beyond the direct income effects.

Second, the paper has remained relatively silent on the impacts on firms of large-scale shifts in the demographics of their consumer base and their responses. For example, what are the characteristics of firms that thrive and that wither in their market presence, in terms of R&D intensity, capital intensity, skill intensity, and product mix? It would be interesting to connect the consumer panel data used in this paper with the firm panel data, to full characterise the interaction of supply- and demand-side mechanisms, and to quantify the welfare effects allowing for endogenous income, given exogenous shocks to the supply- or demand-side parameters. We leave these extensions to future research.

References

Argente, D., Lee, M., 2021. Cost of living inequality during the great recession. Journal of the European Economic Association 19, 913–952.

Atkin, D., Faber, B., Gonzalez-Navarro, M., 2018. Retail globalization and household welfare: Evidence from Mexico. Journal of Political Economy 126, 1–73.

Autor, D., 2019. Work of the past, work of the future. AEA Papers and Proceedings 109, 1–32.

Autor, D., Dorn, D., 2013. The growth of low-skill service jobs and the polarization of the US labor market. American Economic Review 103, 1553–1597.

Bloomberg, 2021a. Millions are tumbling out of the global middle class in historic setback. Bloomberg Businessweek. April 07, 2021.

—, 2021b. The rich are minting money in the pandemic like never before. Bloomberg Wealth. Jan 17, 2021.

Jaravel, X., 2019. The unequal gains from product innovations: Evidence from the U.S. retail sector. Quarterly Journal of Economics 134, 715–783.

Kochhar, R., Fry, R., 2015. The american middle class is losing ground: No longer the majority and falling behind financially. Tech. Rep., Pew Research Center.

Krueger, A., 2012. The rise and consequences of inequality in the United States. Remarks by Chairman Alan B. Krueger of the Council of Economic Advisers at the Center for American Progress, Washington, D.C.

Melitz, M. J., 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. Econometrica 71, 1695–1725.

New York Times, 2014a. Income gap changing US consumer landscape. The New York Times. Feb 04, 2014.

—, 2014b. Mid-tier stores close while posh ones thrive. The New York Times. Feb 04, 2014.

Tayor, P., et al., 2008. Inside the middle class: Bad times hit the good life. Tech. Rep., Pew Research Center.

A Data Appendix

A.1 Processing of the Nielsen Consumer Panel data (2005–2017)

The Nielsen Consumer Panel data (2005–2017) contains seven types of files. Amongst them, we mainly use the files on: panelists, trips, purchases, and products.

The file on panelists reports the household-level demographics information. We use the household income and household size information to compute a proxy of per-capita household income. Note that the Nielsen Consumer Panel data do not report the exact amount of household income, but only the band in which the household income falls. We proxy the household income by the lower bound of the income band that the household belongs to. As such, the actual household income is no smaller than the proxy income. We convert the geographical location of the household reported at the FIPS county code level to the commuting-zone level, using a crosswalk provided by Autor and Dorn (2013).

The file on trips contains information that links each trip to the individual households. We use it in conjunction with the file on purchases to link bar-code level purchases to each household.

The file on purchases contains detailed information on what the purchase basket is for each trip. The information that we use from this file is the trip code, the UPC code, the quantity purchased, and the "final price," which is the total amount (after coupons etc.) paid by the consumers for the UPC. As this file does not contain any household information, we merge it with the file on trips to link trips to households and to aggregate these trips to an annual consumption measure at the household-barcode level.

The file on products provides information for each product at the bar-code level. It provides the product module code (PMC), product group code (PGC), number of units in a multipack (multi), the numeric size of the product (size1_amount), the unit of measure for the amount (size1_unit). We use these information to compute the bar-code level volume that each household purchases following an identical procedure for the retail scanner data (quantity*size1_amount*multi). Bar-code level prices are computed using total expenditure ("final price") divided by total volume for each bar-code product (i.e., each UPC) annually.

B Appendix to Section 3

In this appendix, we report the result for the population share of L/M/H-income groups based on adjusted per-capita income, defined as in Kochhar and Fry (2015) of Pew

Research Center:

$$\text{Adjusted per-capita income} = \frac{\text{Gross household income}}{\text{Number of persons}^N}, \quad 0 \leq N \leq 1, \tag{A1}$$

where N is inversely proportional to the strength of within-household "economies of scale." If $N = 1$, the adjusted income per capita in equation (A1) is equivalent to simple per-capita income, and within-household "economies of scale" is absent. If N is set to 0, the adjusted income per capita is equivalent to the gross household income; in this case, the household income is not discounted by family size. We follow the literature's convention and set $N = 0.5$.[9]

Based on this alternative definition of per-capita income, the time-series of the population share of different income classes at the U.S. national, state, and commuting-zone level are reported in Figures A.1–A.3. The patterns are similar to the benchmark qualitatively, although quantitatively the current set of statistics contain slightly more variations.

[Insert Figures A.1–A.3 Here]

C Appendix to Section 4

This appendix reports the estimation results of equations (3) and (4) when the population share of L/M/H-income class is calculated based on the distribution of adjusted per-capita income, where the adjusted per-capita income is measured according to equation (A1). The findings are largely similar to the benchmark, but with weaker (or insignificant) effects for the L-priced products (L-income class) at the state level. Second, the effect of L-income population shares on L-priced product market shares is still positive in less densely populated commuting zones (across the three sets of price cutoffs), but the signs of the effects are not robust in more densely populated areas. They are negative with tercile and quartile price cutoffs, but positive with decile price cutoffs. The interpretation is similar to the benchmark (but with some modifications) that although the population shares of L-income workers increases in general over the years (at the extensive margin), their wages might not have improved *or even worsened* (at the intensive margin) in the more densely populated areas, such that *their purchasing powers weaken*, which was manifested in an increased demand for extremely cheap products (under the 10th percentile) but *possibly much more reduced demand* for products priced in between the 10th and 33rd percentiles. Full details of the estimation results can be found in Tables A.1–A.9.

[9] The economies of scale may be present within household in the sense that goods are cheaper when purchased in bulks. For example, a one-bedroom apartment is usually more costly to rent than a two-bedroom apartment in terms of cost per bedroom due to the sharing of public space such as living room and pantry. Similarly, a larger household enjoys larger discounts when purchasing groceries/consumables in larger quantities.

[Insert Tables A.1–A.9 Here]

Table 1: Summary Statistics of the IPUMS ACS Sample (2005–2017)

Year	2005	2017
Number of households	1,142,884	1,220,348
Number of unique households across years (2005–2017)		2,739,609
Average number of persons in each household	2.49	2.45
25% unadjusted household income per capita	8,957	9,336
50% unadjusted household income per capita	15,968	16,975
75% unadjusted household income per capita	26,806	29,197
25% adjusted household income per capita	15,781	16,327
50% adjusted household income per capita	27,579	29,106
75% adjusted household income per capita	44,031	47,628

Notes: We have trimmed the raw sample as follows: (1) dropping group quarters ($\approx$ 180,000 unique households across years); (2) dropping observations if household income == 9999999 (NA); conditional on (1), (2) drops nothing; (3) dropping observations if household income is negative ($\approx$ 29,000 unique households across years); (4) dropping observations if household income is missing (==.); conditional on (1)–(3), (4) drops nothing; (5) dropping observations of Alaska and Hawaii ($\approx$ 16,000 unique households across years); (6) dropping observations if census tract identifier is missing; conditional on (1)–(5), (6) drops nothing. The income is in 1999 real dollars.

Table 2: Summary Statistics of the IPUMS ACS Sample by Population Density of Commuting Zones in 2005

Bin	Unadjusted income	Adjusted income	% Lower	% Middle	% Upper	Density	No. of c-zones
1	14,501	21,412	39.9	52.1	8.0	18	398
2	14,757	21,895	38.8	52.1	9.1	70	140
3	15,823	23,884	34.9	54.2	10.9	115	66
4	16,642	24,850	32.8	55.2	11.9	181	44
5	18,212	27,143	29.2	55.1	15.7	269	26
6	18,766	28,107	27.6	55.7	16.7	349	18
7	19,193	29,434	29.0	52.7	18.2	464	8
8	20,557	31,220	26.5	52.7	20.8	710	9
9	21,752	33,085	23.5	52.4	24.1	1,004	8
10	22,249	34,277	23.1	51.2	25.7	2,302	5

Notes: Columns 2–3 report the average income across commuting zones within the same bin of population density. Columns 4–6 report the percentage of the population within each bin that belong to the lower/middle/upper income class, where the cutoffs for the middle income class are 2/3 and twice of the national median income level. Density is defined as the number of residents per squared miles.

Table 3: Product Market Shares on Lagged Population Shares — State-Level; Tercile Price Cutoffs

	$Share_{s,t,g,L}$ (1)	$Share_{s,t,g,M}$ (2)	$Share_{s,t,g,H}$ (3)	$Share_{s,t,g,L}$ (4)	$Share_{s,t,g,M}$ (5)	$Share_{s,t,g,H}$ (6)
$Frac_{s,t-1,L}$	0.181***			0.173***		
	(0.049)			(0.039)		
$Frac_{s,t-1,M}$		0.362***			0.362***	
		(0.048)			(0.042)	
$Frac_{s,t-1,H}$			0.721***			0.694***
			(0.053)			(0.041)
Robust SE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Product FE	No	No	No	Yes	Yes	Yes
R^2	0.016	0.005	0.014	0.441	0.299	0.514
N	57,218	57,218	57,218	57,218	57,218	57,218

Notes: Market share measures are constructed using the Nielsen Retail Scanner Data (2006–2017) as documented in Section 2.1. Population share measures are constructed using the IPUMS ACS Data (2005–2017) as documented in Section 3.1. States include all the contiguous states in the United States and the District of Columbia. All regressions include a constant term. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. In this exercise, the L-priced (H-priced) products are defined as those UPCs whose deflated real price is below the 33rd (above the 66th) percentile of the year-2006 price distribution (of the corresponding PGC g in area s). The M-priced products are defined as those with real prices in between the two terciles. Income per capita is calculated based on formula (2).

Table 4: Product Market Shares on Lagged Population Shares — State-Level; Quartile Price Cutoffs

	$Share_{s,t,g,L}$ (1)	$Share_{s,t,g,M}$ (2)	$Share_{s,t,g,H}$ (3)	$Share_{s,t,g,L}$ (4)	$Share_{s,t,g,M}$ (5)	$Share_{s,t,g,H}$ (6)
$Frac_{s,t-1,L}$	0.101**			0.097***		
	(0.044)			(0.036)		
$Frac_{s,t-1,M}$		0.481***			0.484***	
		(0.050)			(0.043)	
$Frac_{s,t-1,H}$			0.746***			0.720***
			(0.050)			(0.039)
Robust SE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Product FE	No	No	No	Yes	Yes	Yes
R^2	0.014	0.005	0.016	0.430	0.349	0.532
N	57,218	57,218	57,218	57,218	57,218	57,218

Notes: See the footnote of Table 3. In this exercise, the L-priced (H-priced) products are defined as those UPCs whose deflated real price is below the 25th (above the 75th) percentile of the year-2006 price distribution (of the corresponding PGC g in area s). The M-priced products are defined as those with real prices in between the two quartiles. Income per capita is calculated based on formula (2).

Table 5: Product Market Shares on Lagged Population Shares — State-Level; Decile Price Cutoffs

	$Share_{s,t,g,L}$ (1)	$Share_{s,t,g,M}$ (2)	$Share_{s,t,g,H}$ (3)	$Share_{s,t,g,L}$ (4)	$Share_{s,t,g,M}$ (5)	$Share_{s,t,g,H}$ (6)
$Frac_{s,t-1,L}$	0.080**			0.079***		
	(0.031)			(0.027)		
$Frac_{s,t-1,M}$		0.445***			0.448***	
		(0.042)			(0.036)	
$Frac_{s,t-1,H}$			0.549***			0.532***
			(0.035)			(0.027)
Robust SE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Product FE	No	No	No	Yes	Yes	Yes
R^2	0.008	0.005	0.015	0.366	0.376	0.539
N	57,218	57,218	57,218	57,218	57,218	57,218

Notes: See the footnote of Table 3. In this exercise, the L-priced (H-priced) products are defined as those UPCs whose deflated real price is below the 10th (above the 90th) percentile of the year-2006 price distribution (of the corresponding PGC g in area s). The M-priced products are defined as those with real prices in between the two deciles. Income per capita is calculated based on formula (2).

Table 6: Product Market Shares on Lagged Population Shares — Commuting-Zone-Level; Tercile Price Cutoffs

	$Share_{s,t,g,L}$ (1)	$Share_{s,t,g,M}$ (2)	$Share_{s,t,g,H}$ (3)	$Share_{s,t,g,L}$ (4)	$Share_{s,t,g,M}$ (5)	$Share_{s,t,g,H}$ (6)
$Frac_{s,t-1,L}$	0.149***			0.136***		
	(0.010)			(0.008)		
$Frac_{s,t-1,M}$		0.087***			0.092***	
		(0.010)			(0.009)	
$Frac_{s,t-1,H}$			0.556***			0.555***
			(0.015)			(0.012)
Robust SE	Yes	Yes	Yes	Yes	Yes	Yes
CZone FE	Yes	Yes	Yes	Yes	Yes	Yes
Product FE	No	No	No	Yes	Yes	Yes
R^2	0.234	0.077	0.103	0.491	0.275	0.466
N	642,031	642,031	642,031	642,031	642,031	642,031

Notes: Market share measures are constructed using the Nielsen Retail Scanner Data (2006–2017) as documented in Section 2.1. Population share measures are constructed using the IPUMS ACS Data (2005–2017) as documented in Section 3.1. Note that the Nielsen Retail Scanner Data do not report sales data for all 722 commuting zones in the contiguous U.S., especially in those least populated areas. In particular, the Nielsen Retail Scanner Data cover 660 commuting zones in 2006 (min), 690 in 2016 (max), and 688 in 2017. All of the missing commuting zones belong to Bin 1 in Table 2, except one that belongs to Bin 2. All regressions include a constant term. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. In this exercise, the L-priced (H-priced) products are defined as those UPCs whose deflated real price is below the 33rd (above the 66th) percentile of the year-2006 price distribution (of the corresponding PGC g in area s). The M-priced products are defined as those with real prices in between the two terciles. Income per capita is calculated based on formula (2).

Table 7: Product Market Shares on Lagged Population Shares — Commuting-Zone-Level; Quartile Price Cutoffs

	$Share_{s,t,g,L}$ (1)	$Share_{s,t,g,M}$ (2)	$Share_{s,t,g,H}$ (3)	$Share_{s,t,g,L}$ (4)	$Share_{s,t,g,M}$ (5)	$Share_{s,t,g,H}$ (6)
$Frac_{s,t-1,L}$	0.133***			0.121***		
	(0.009)			(0.008)		
$Frac_{s,t-1,M}$		0.115***			0.122***	
		(0.011)			(0.009)	
$Frac_{s,t-1,H}$			0.569***			0.573***
			(0.014)			(0.011)
Robust SE	Yes	Yes	Yes	Yes	Yes	Yes
CZone FE	Yes	Yes	Yes	Yes	Yes	Yes
Product FE	No	No	No	Yes	Yes	Yes
R^2	0.290	0.131	0.083	0.513	0.365	0.472
N	642,031	642,031	642,031	642,031	642,031	642,031

Notes: See the footnote of Table 6. In this exercise, the L-priced (H-priced) products are defined as those UPCs whose deflated real price is below the 25th (above the 75th) percentile of the year-2006 price distribution (of the corresponding PGC g in area s). The M-priced products are defined as those with real prices in between the two quartiles. Income per capita is calculated based on formula (2).

Table 8: Product Market Shares on Lagged Population Shares — Commuting-Zone-Level; Decile Price Cutoffs

	$Share_{s,t,g,L}$ (1)	$Share_{s,t,g,M}$ (2)	$Share_{s,t,g,H}$ (3)	$Share_{s,t,g,L}$ (4)	$Share_{s,t,g,M}$ (5)	$Share_{s,t,g,H}$ (6)
$Frac_{s,t-1,L}$	0.063***			0.056***		
	(0.007)			(0.007)		
$Frac_{s,t-1,M}$		0.111***			0.124***	
		(0.010)			(0.009)	
$Frac_{s,t-1,H}$			0.424***			0.437***
			(0.011)			(0.009)
Robust SE	Yes	Yes	Yes	Yes	Yes	Yes
CZone FE	Yes	Yes	Yes	Yes	Yes	Yes
Product FE	No	No	No	Yes	Yes	Yes
R^2	0.453	0.296	0.058	0.577	0.530	0.521
N	642,031	642,031	642,031	642,031	642,031	642,031

Notes: See the footnote of Table 6. In this exercise, the L-priced (H-priced) products are defined as those UPCs whose deflated real price is below the 10th (above the 90th) percentile of the year-2006 price distribution (of the corresponding PGC g in area s). The M-priced products are defined as those with real prices in between the two deciles. Income per capita is calculated based on formula (2).

Table 9: Product Market Shares on Lagged Population Shares ($\times\mathbf{1}$\{Population Density $\geq$ National Median\}) — Commuting-Zone-Level; Tercile Price Cutoffs

	$Share_{s,t,g,L}$ (1)	$Share_{s,t,g,M}$ (2)	$Share_{s,t,g,H}$ (3)	$Share_{s,t,g,L}$ (4)	$Share_{s,t,g,M}$ (5)	$Share_{s,t,g,H}$ (6)
$Frac_{s,t-1,L}$	0.155***			0.143***		
	(0.010)			(0.008)		
$Frac_{s,t-1,L}$	-0.110**			-0.116***		
$\times\mathbf{1}\{Density_s \geq median\}$	(0.055)			(0.043)		
$Frac_{s,t-1,M}$		0.067***			0.072***	
		(0.010)			(0.009)	
$Frac_{s,t-1,M}$		0.477***			0.478***	
$\times\mathbf{1}\{Density_s \geq median\}$		(0.039)			(0.035)	
$Frac_{s,t-1,H}$			0.541***			0.540***
			(0.016)			(0.013)
$Frac_{s,t-1,H}$			0.921***			0.911***
$\times\mathbf{1}\{Density_s \geq median\}$			(0.071)			(0.053)
Robust SE	Yes	Yes	Yes	Yes	Yes	Yes
CZone FE	Yes	Yes	Yes	Yes	Yes	Yes
Product FE	No	No	No	Yes	Yes	Yes
R^2	0.234	0.077	0.103	0.491	0.275	0.466
N	642,031	642,031	642,031	642,031	642,031	642,031

Notes: See the footnote of Table 6. We compute the population density for each CZ, using a crosswalk from the county level to the CZ level provided by Autor and Dorn (2013), the county population in 2005 provided by the NBER, and the county land areas provided by the Census Bureau. The indicator $\mathbf{1}\{Density_s \geq median\}$ equals one if the population density of CZ s is above or equal to the national median (across all 722 commuting zones). In this exercise, the L-priced (H-priced) products are defined as those UPCs whose deflated real price is below the 33rd (above the 66th) percentile of the year-2006 price distribution (of the corresponding PGC g in area s). The M-priced products are defined as those with real prices in between the two terciles. Income per capita is calculated based on formula (2).

Table 10: Product Market Shares on Lagged Population Shares ($\times\mathbf{1}\{$Population Density $\geq$ National Median$\}$) — Commuting-Zone-Level; Quartile Price Cutoffs

	$Share_{s,t,g,L}$ (1)	$Share_{s,t,g,M}$ (2)	$Share_{s,t,g,H}$ (3)	$Share_{s,t,g,L}$ (4)	$Share_{s,t,g,M}$ (5)	$Share_{s,t,g,H}$ (6)
$Frac_{s,t-1,L}$	0.138***			0.125***		
	(0.009)			(0.008)		
$Frac_{s,t-1,L}$	-0.043			-0.048		
$\times\mathbf{1}\{Density_s \geq median\}$	(0.051)			(0.040)		
$Frac_{s,t-1,M}$		0.088***			0.095***	
		(0.011)			(0.010)	
$Frac_{s,t-1,M}$		0.643***			0.646***	
$\times\mathbf{1}\{Density_s \geq median\}$		(0.043)			(0.037)	
$Frac_{s,t-1,H}$			0.552***			0.556***
			(0.015)			(0.012)
$Frac_{s,t-1,H}$			0.973***			0.965***
$\times\mathbf{1}\{Density_s \geq median\}$			(0.067)			(0.049)
Robust SE	Yes	Yes	Yes	Yes	Yes	Yes
CZone FE	Yes	Yes	Yes	Yes	Yes	Yes
Product FE	No	No	No	Yes	Yes	Yes
R^2	0.290	0.132	0.083	0.514	0.365	0.473
N	642,031	642,031	642,031	642,031	642,031	642,031

Notes: See the footnote of Table 9. In this exercise, the L-priced (H-priced) products are defined as those UPCs whose deflated real price is below the 25th (above the 75th) percentile of the year-2006 price distribution (of the corresponding PGC g in area s). The M-priced products are defined as those with real prices in between the two quartiles. Income per capita is calculated based on formula (2).

Table 11: Product Market Shares on Lagged Population Shares ($\times\mathbf{1}\{$Population Density $\geq$ National Median$\}$) — Commuting-Zone-Level; Decile Price Cutoffs

	$Share_{s,t,g,L}$	$Share_{s,t,g,M}$	$Share_{s,t,g,H}$	$Share_{s,t,g,L}$	$Share_{s,t,g,M}$	$Share_{s,t,g,H}$
	(1)	(2)	(3)	(4)	(5)	(6)
$Frac_{s,t-1,L}$	0.062***			0.054***		
	(0.008)			(0.007)		
$Frac_{s,t-1,L}$	0.132***			0.125***		
$\times\mathbf{1}\{Density_s \geq median\}$	(0.039)			(0.034)		
$Frac_{s,t-1,M}$		0.085***			0.098***	
		(0.011)			(0.009)	
$Frac_{s,t-1,M}$		0.612***			0.631***	
$\times\mathbf{1}\{Density_s \geq median\}$		(0.044)			(0.032)	
$Frac_{s,t-1,H}$			0.415***			0.428***
			(0.011)			(0.009)
$Frac_{s,t-1,H}$			0.629***			0.655***
$\times\mathbf{1}\{Density_s \geq median\}$			(0.056)			(0.032)
Robust SE	Yes	Yes	Yes	Yes	Yes	Yes
CZone FE	Yes	Yes	Yes	Yes	Yes	Yes
Product FE	No	No	No	Yes	Yes	Yes
R^2	0.453	0.296	0.058	0.577	0.530	0.521
N	642,031	642,031	642,031	642,031	642,031	642,031

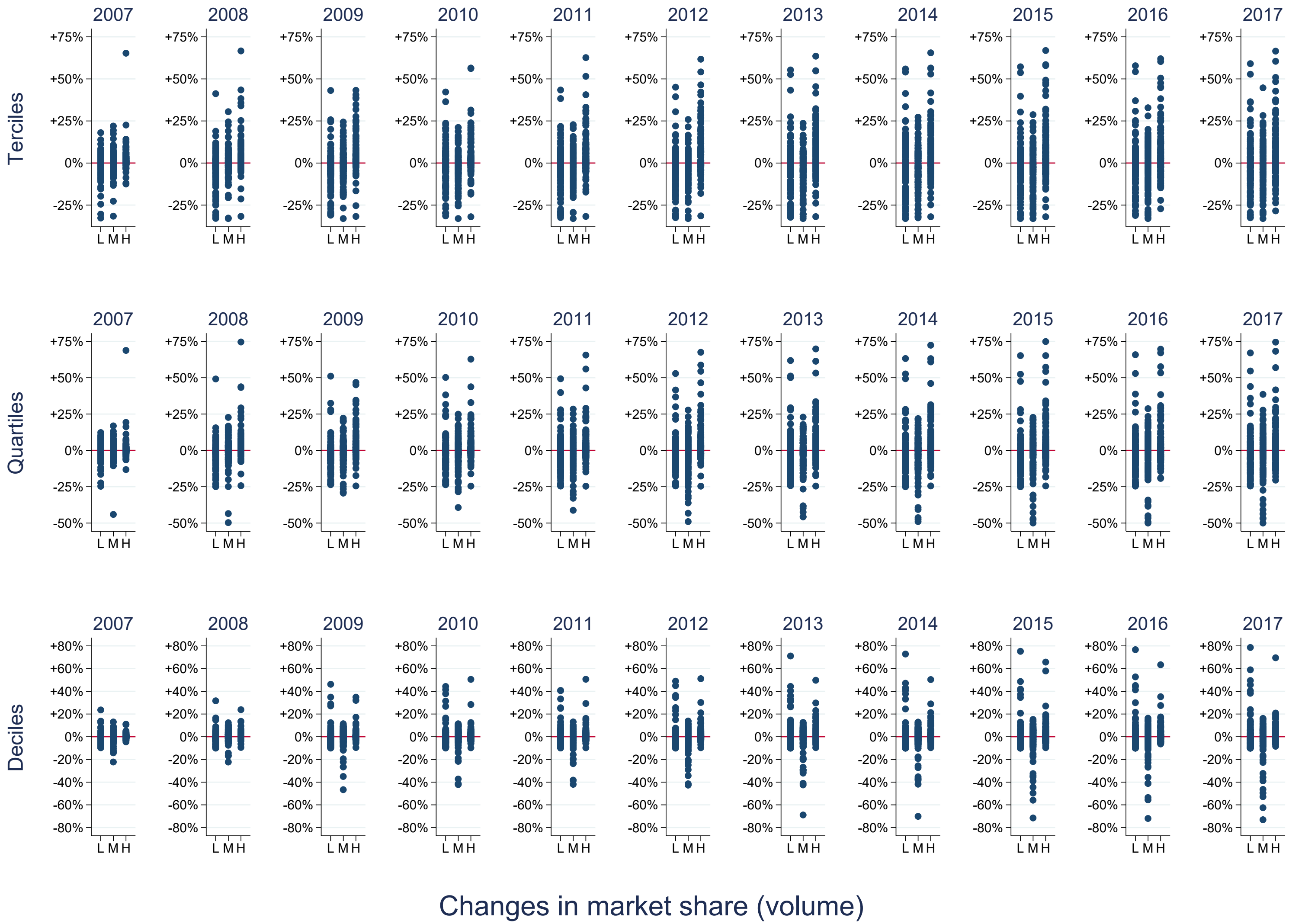
Notes: See the footnote of Table 9. In this exercise, the L-priced (H-priced) products are defined as those UPCs whose deflated real price is below the 10th (above the 90th) percentile of the year-2006 price distribution (of the corresponding PGC g in area s). The M-priced products are defined as those with real prices in between the two deciles. Income per capita is calculated based on formula (2).

Table 12: Summary Statistics of the Nielsen Consumer Panel data (2005–2017)

Year	2005	2017
Number of households	38,862	62,829
Number of unique households across years (2005–2017)		171,690
Average number of persons in each household	2.34	2.44
Share of lower-income households	25.2%	29.1%
Share of middle-income households	61.6%	54.1%
Share of upper-income households	13.2%	16.8%
Per-capita consumption of lower-income households	1,322.8	1,988.6
Per-capita consumption of middle-income households	1,702.6	2,581.7
Per-capita consumption of upper-income households	2,242.4	3,373.8

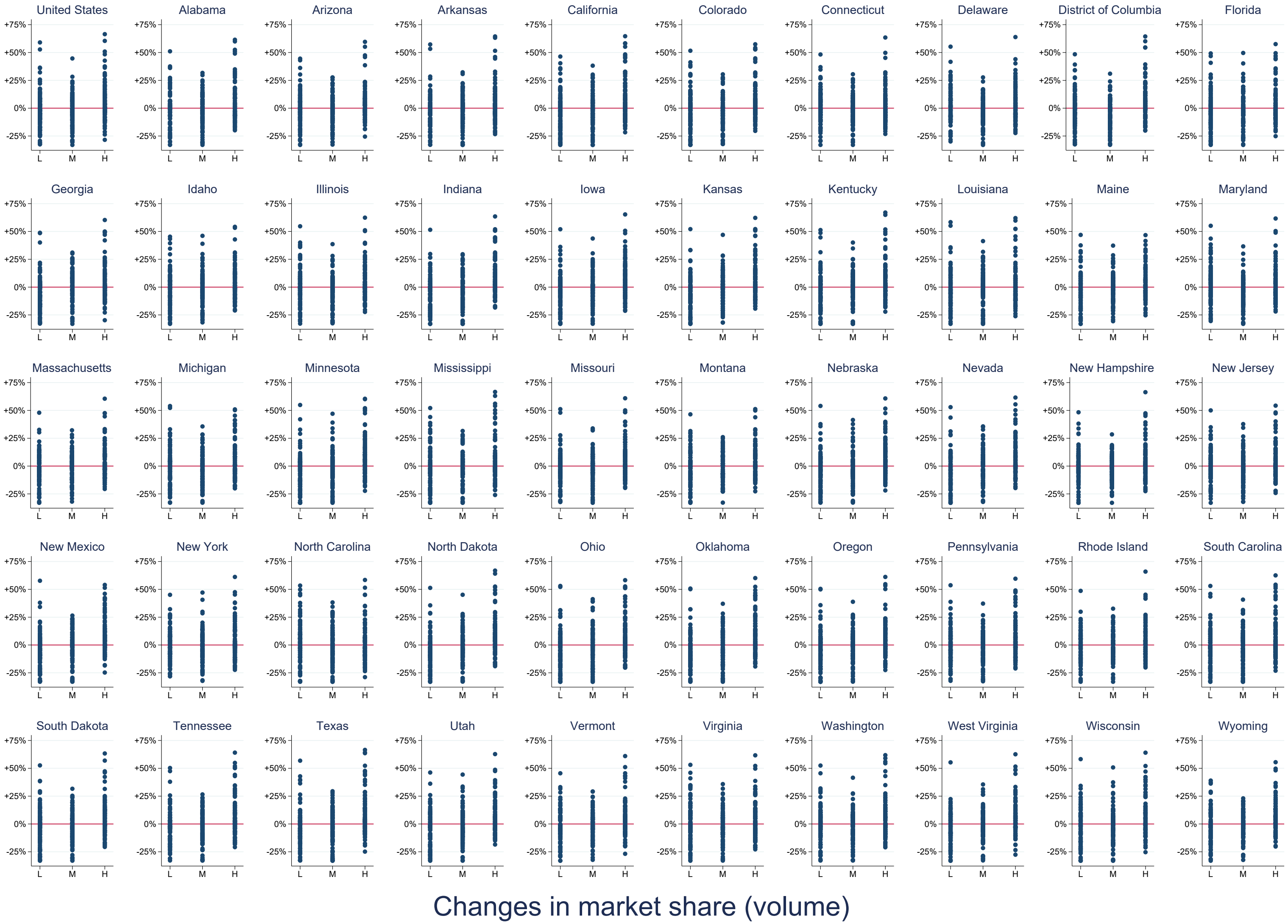
Notes: Note that the Nielsen Consumer Panel data do not report the exact amount of household income, but only the band in which the household income falls. We proxy the household income by the lower bound of the income band that the household belongs to. As such, the actual household income is no smaller than the proxy income. We use the household income and household size information to compute a proxy of per-capita household income. The income cutoffs for L/M/H-income are based on the statistics of the IPUMS ACS sample of the same year. The consumption value is in current nominal dollars. Further details are provided in Appendix A.1.

Figure 1: Changes in the Market Shares of Low/Middle/High-priced Products across the United States



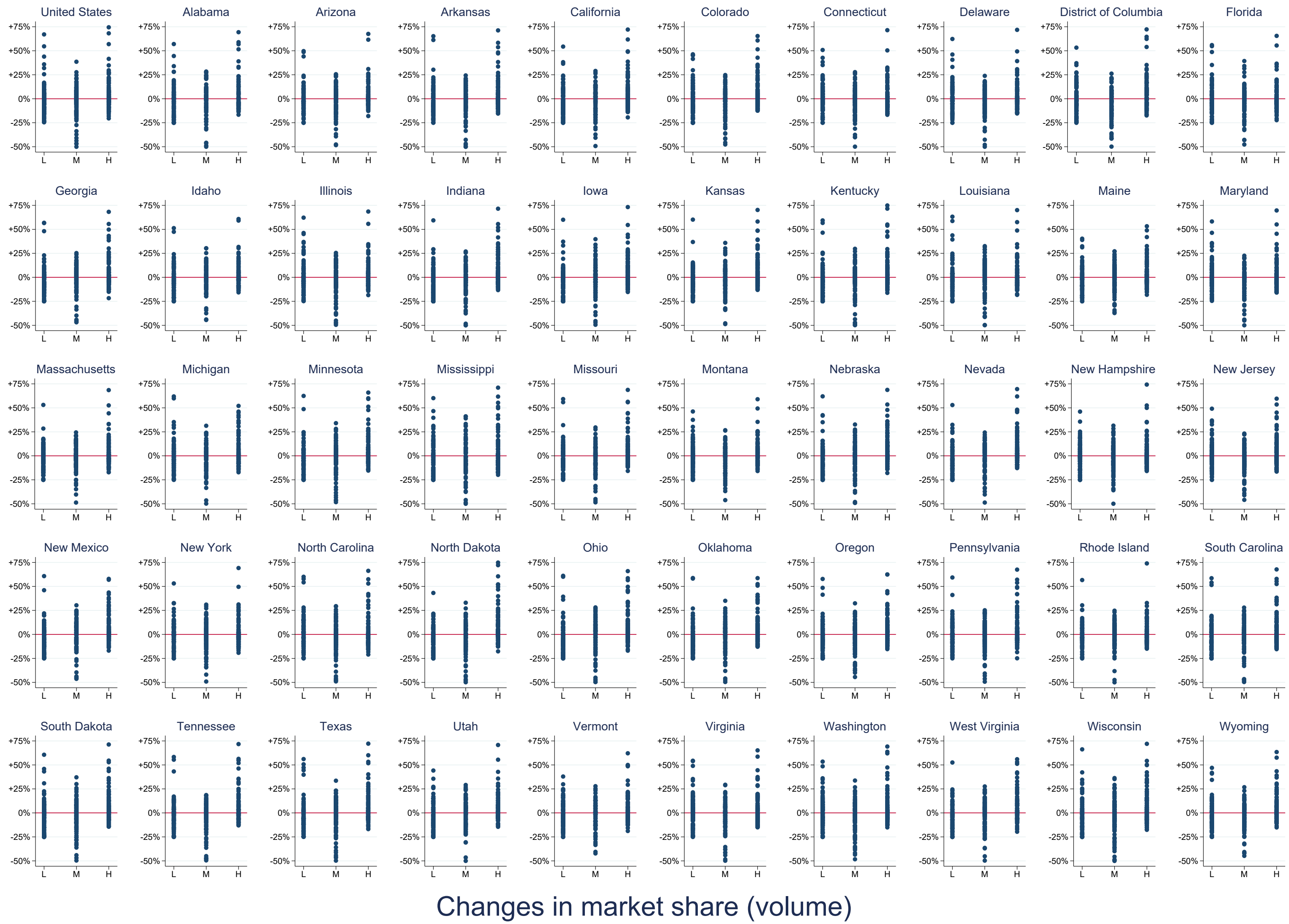
Note: The figure plots the change in the market share (in percentage points) of the low (L), middle (M), and high (H)-priced products for each PGC, in each year relative to the base year 2006, across the U.S. (48 contiguous states and the District of Columbia). In defining the L/M/H-priced products, we have used the lower/upper terciles (T), quartiles (Q), and deciles (D) to define the lower/higher price cutoffs for the middle-priced products, as shown in the first, second, and third rows of the figure, respectively. Data are from Nielsen Retail Scanner Database (2006–2017).

Figure 2: Changes in the Market Shares of Low/Middle/High-priced Products by States (2017 vs. 2006: Tercile Cutoffs)



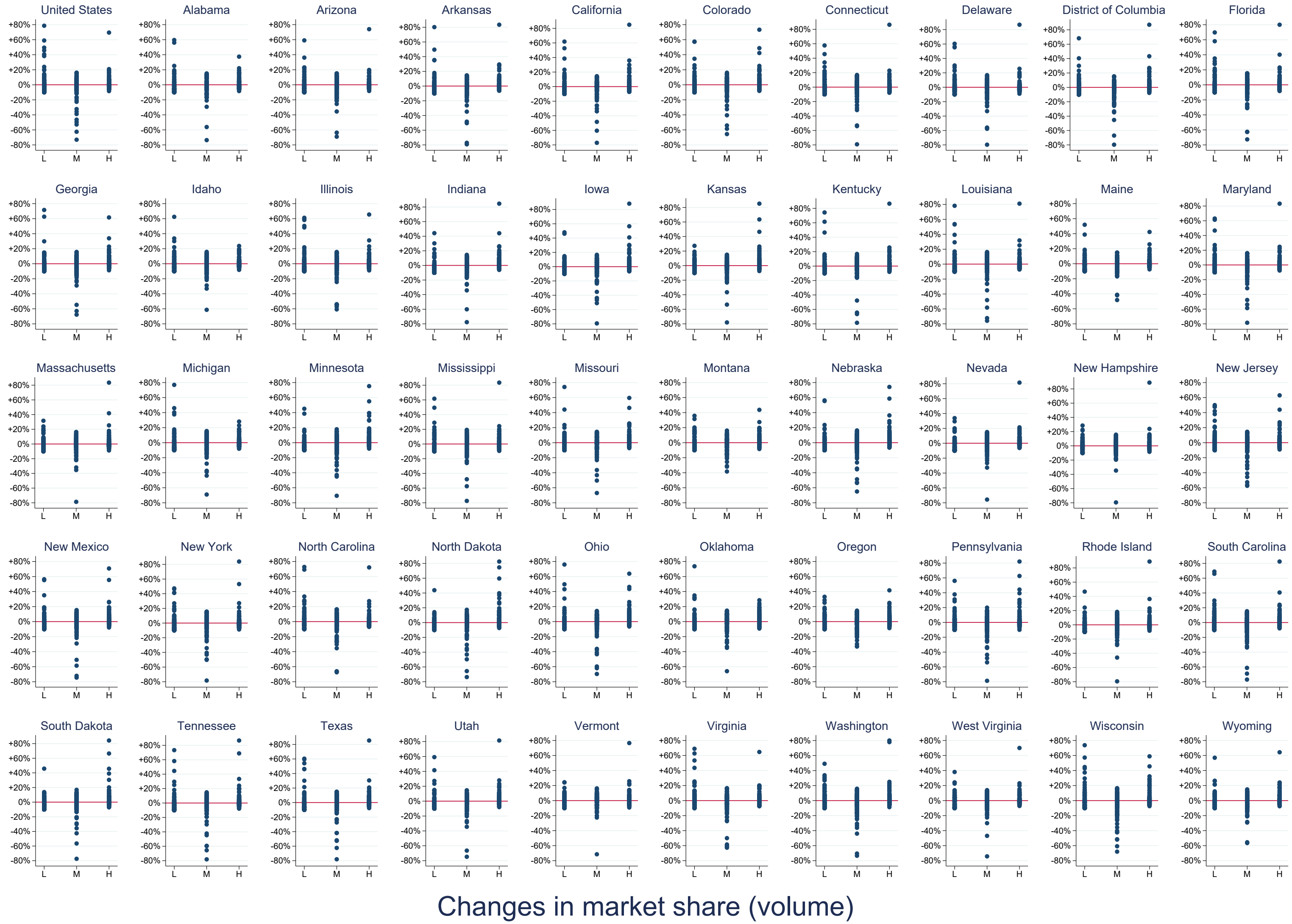
Note: The figure plots the change in the market share (in percentage points) of the low (L), middle (M), and high (H)-priced products for each PGC, in 2017 relative to the base year 2006, in each of the 48 contiguous states and the District of Columbia. In defining the L/M/H-priced products, we have used the lower/upper terciles (T) to define the lower/higher price cutoffs for the middle-priced products. Data are from Nielsen Retail Scanner Database (2006–2017).

Figure 3: Changes in the Market Shares of Low/Middle/High-priced Products by States (2017 vs. 2006; Quartile Cutoffs)



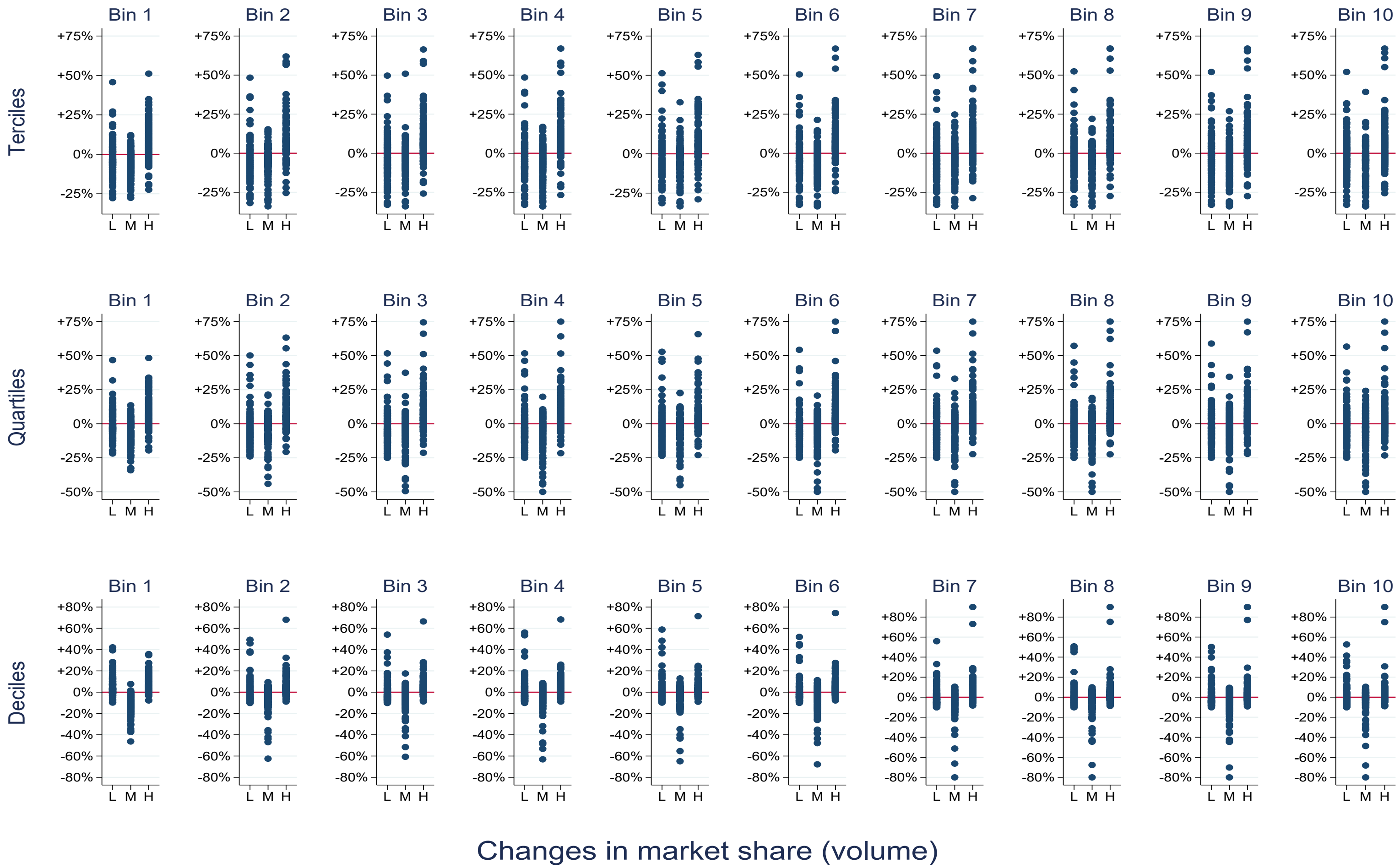
Note: The figure plots the change in the market share (in percentage points) of the low (L), middle (M), and high (H)-priced products for each PGC, in 2017 relative to the base year 2006, in each of the 48 contiguous states and the District of Columbia. In defining the L/M/H-priced products, we have used the lower/upper quartiles (Q) to define the lower/higher price cutoffs for the middle-priced products. Data are from Nielsen Retail Scanner Database (2006–2017).

Figure 4: Changes in the Market Shares of Low/Middle/High-priced Products by States (2017 vs. 2006: Decile Cutoffs)



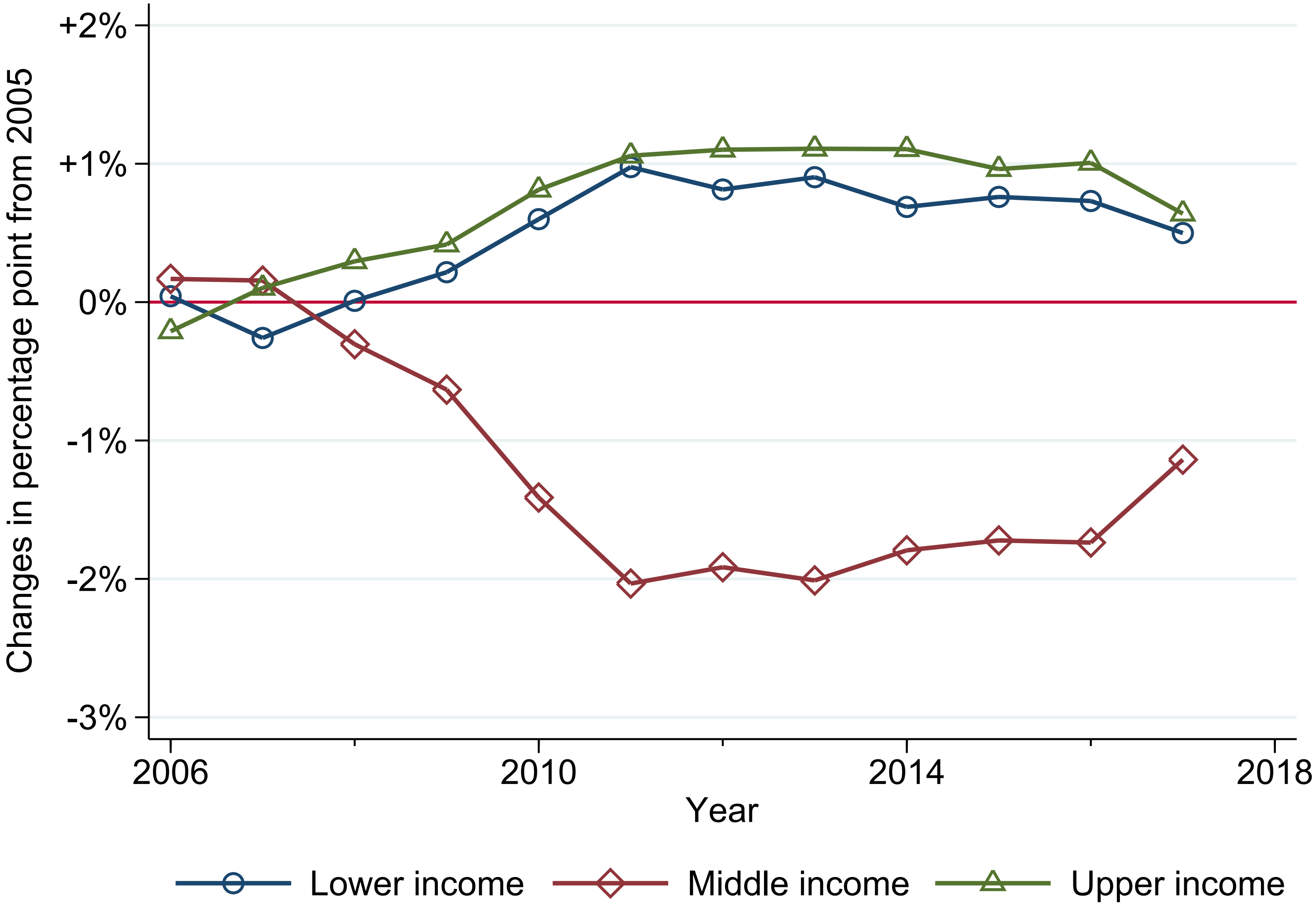
Note: The figure plots the change in the market share (in percentage points) of the low (L), middle (M), and high (H)-priced products for each PGC, in 2017 relative to the base year 2006, in each of the 48 contiguous states and the District of Columbia. In defining the L/M/H-priced products, we have used the lower/upper deciles (D) to define the lower/higher price cutoffs for the middle-priced products. Data are from Nielsen Retail Scanner Database (2006–2017).

Figure 5: Changes in the Market Shares of Low/Middle/High-priced Products by Commuting Zones (2017 vs. 2006)—average across commuting zones within the same bin of population density



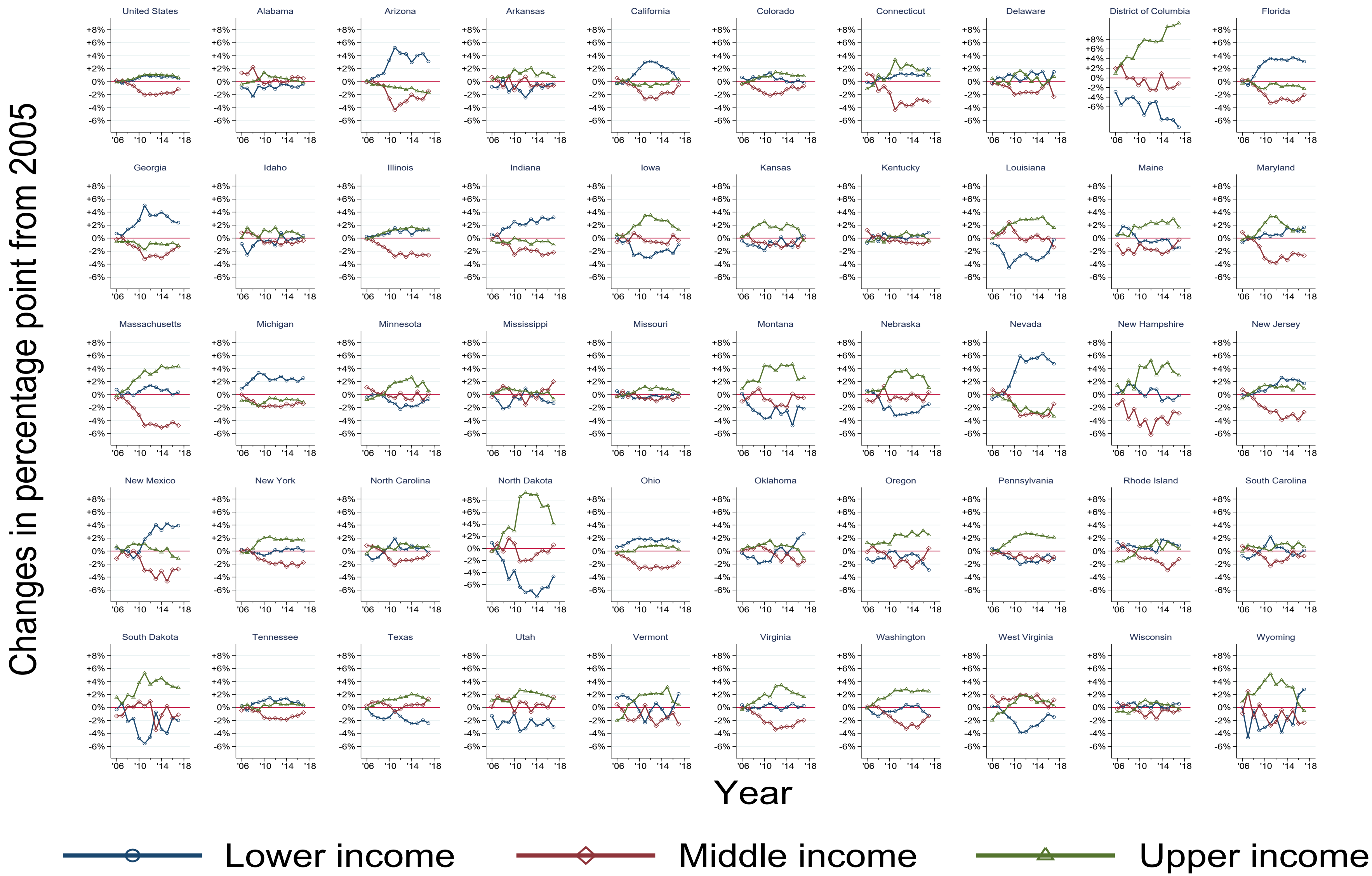
Note: The figure plots the change in the average market share (in percentage points) of the low (L), middle (M), and high (H)-priced products for each PGC, in 2017 relative to the base year 2006, across commuting zones within the same bin. In particular, commuting zones are ranked by their population densities, weighted by the population size of the commuting zone, and grouped into 10 bins. Bin 10 (1) has the highest (lowest) population density. The average market share for each PGC is taken across commuting zones within the same bin of population density. In defining the L/M/H-priced products, we have used the lower/upper terciles (T), quartiles (Q), and deciles (D) to define the lower/higher price cutoffs for the middle-priced products, as shown in the first, second, and third rows of the figure, respectively. Data are from Nielsen Retail Scanner Database (2006–2017).

Figure 6: Changes in Population Shares of Different Income Classes in the United States (relative to 2005)



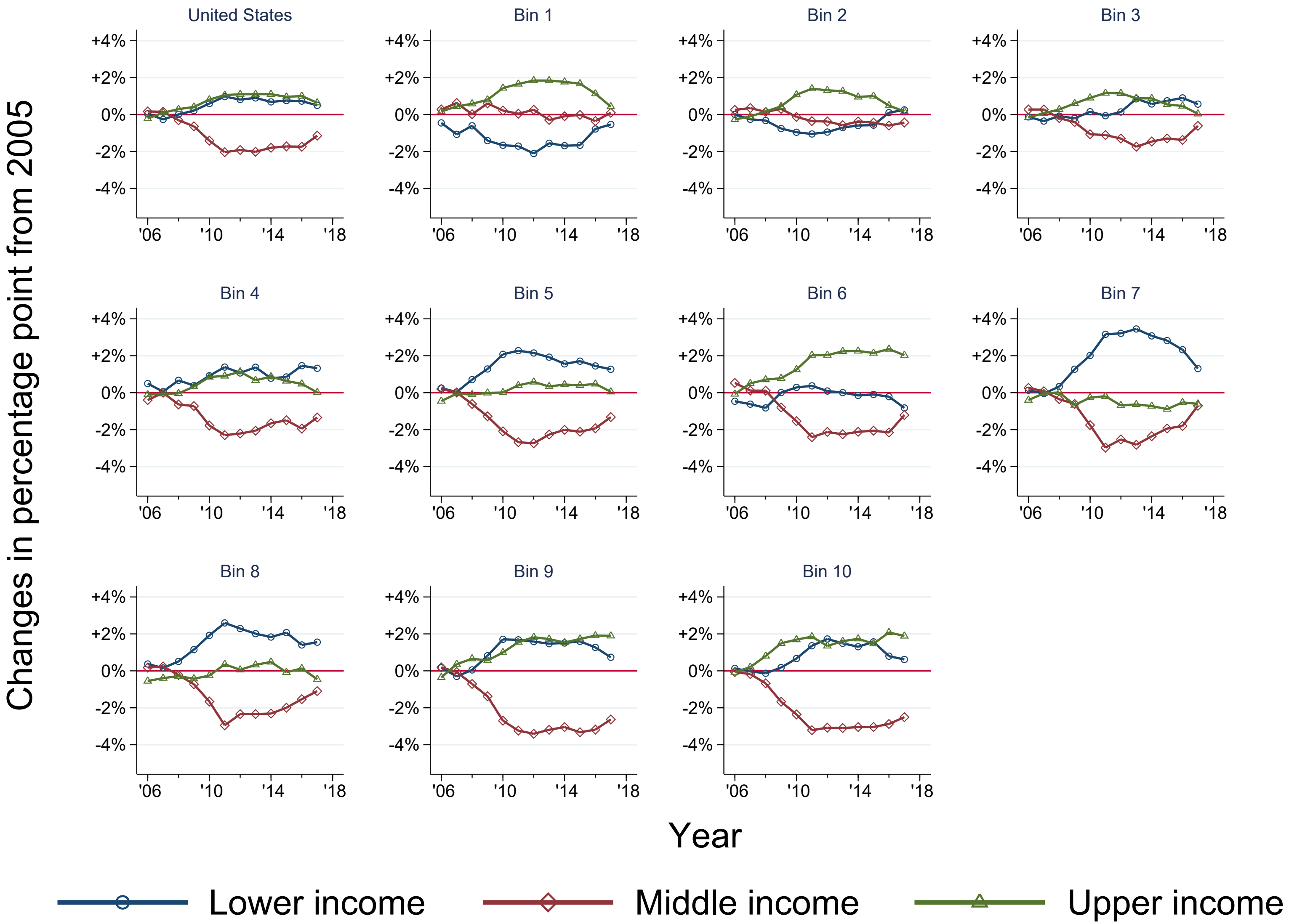
Note: Each plotted point represents the change in the population share (in percentage points) of a particular income group in a year, relative to 2005, for the United States. Income per capita is calculated based on formula (2). Data are from the IPUMS ACS (2005–2017).

Figure 7: Changes in Population Shares of Different Income Classes by States/District



Note: Each plotted point represents the change in the population share (in percentage points) of a particular income group in a year, relative to 2005, for a state/district. Income per capita is calculated based on formula (2). Data are from the IPUMS ACS (2005–2017).

Figure 8: Changes in Population Shares of Different Income Classes by Population Density of Commuting Zones

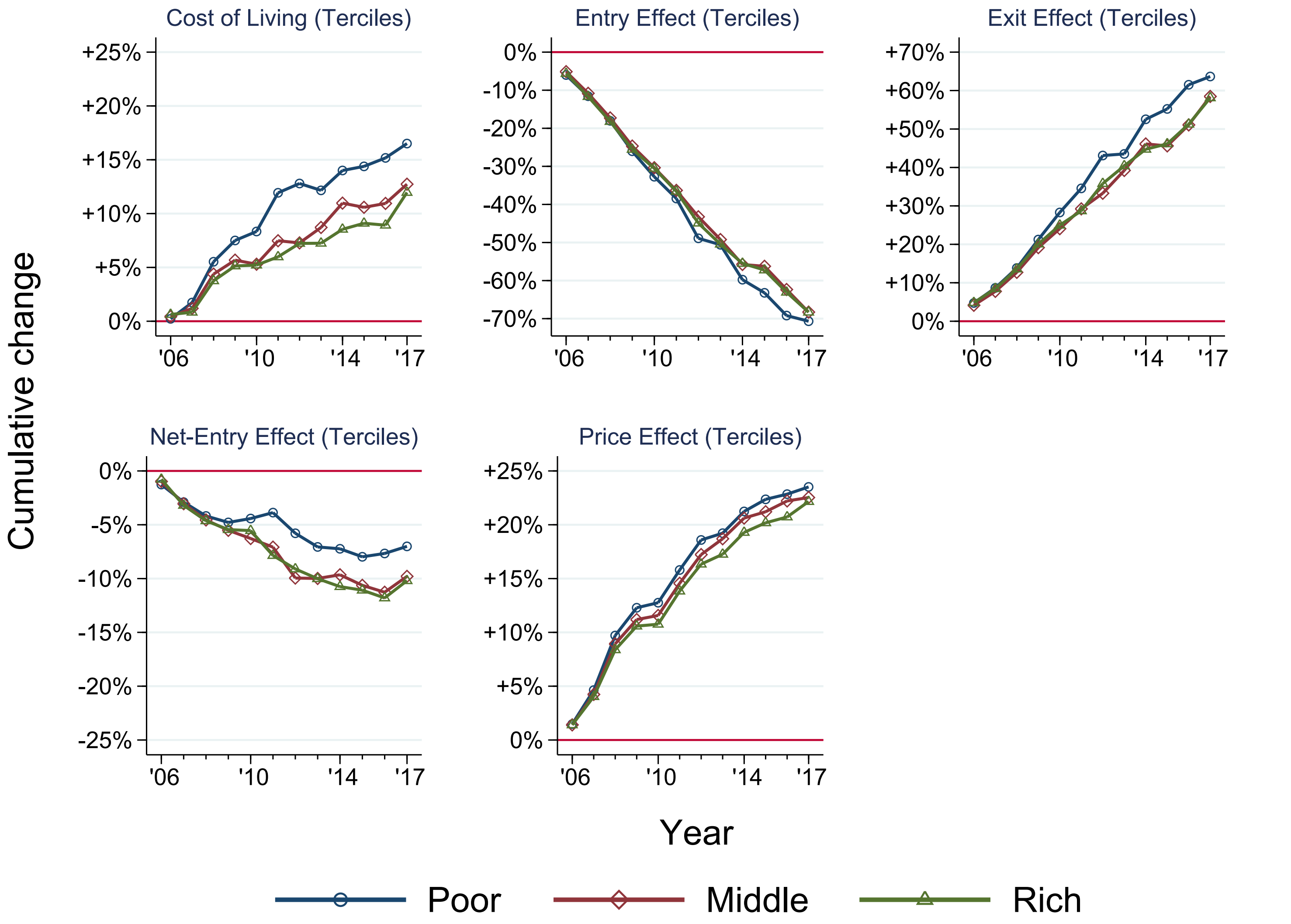


Note: Each plotted point represents the change in the population share (in percentage points) of a particular income group in a year, relative to 2005, for the commuting zones within the same bin of population density. Bin 10 (1) has the highest (lowest) population density. Income per capita is calculated based on formula (2). Data are from the IPUMS ACS (2005–2017).

Figure 9: Non-Homothetic Preferences—Expenditure Shares of L/M/H-Income Groups on L/M/H-Priced Products

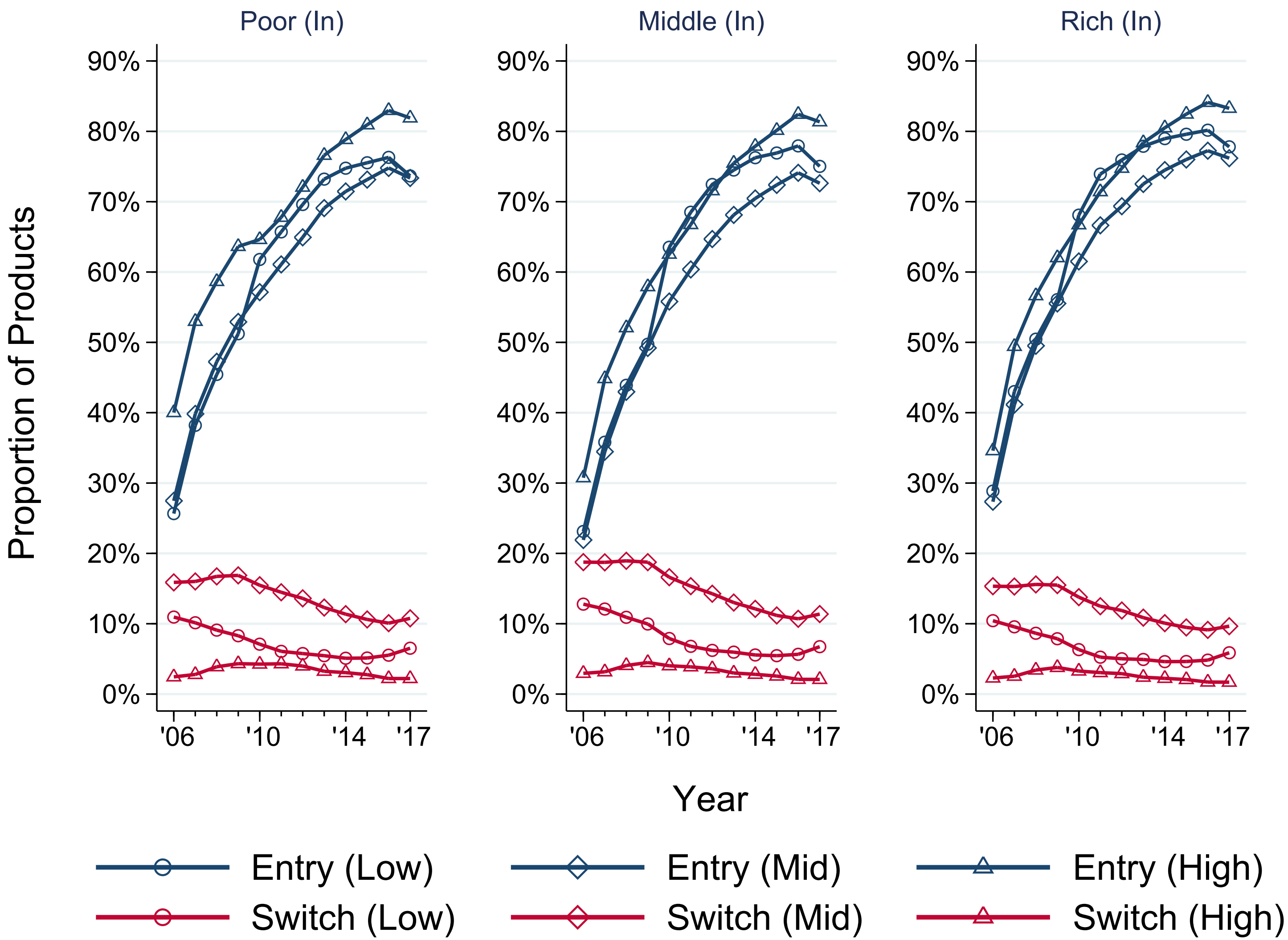
Note: The figure plots the expenditure share on the low (L), middle (M), and high (H)-priced products (defined within each PGC), in each year, by individuals in the L/M/H-income group, respectively, pulling observations across PGCs and the U.S. (48 contiguous states and the District of Columbia). In defining the L/M/H-priced products, we have used the lower/upper terciles (T) of the year-2005 price distributions to define the lower/higher price cutoffs for the middle-priced products. Data are from Nielsen Consumer Panel (2005–2017).

Figure 10: Changes in the Cost of Living and Decomposition (2005–2017)



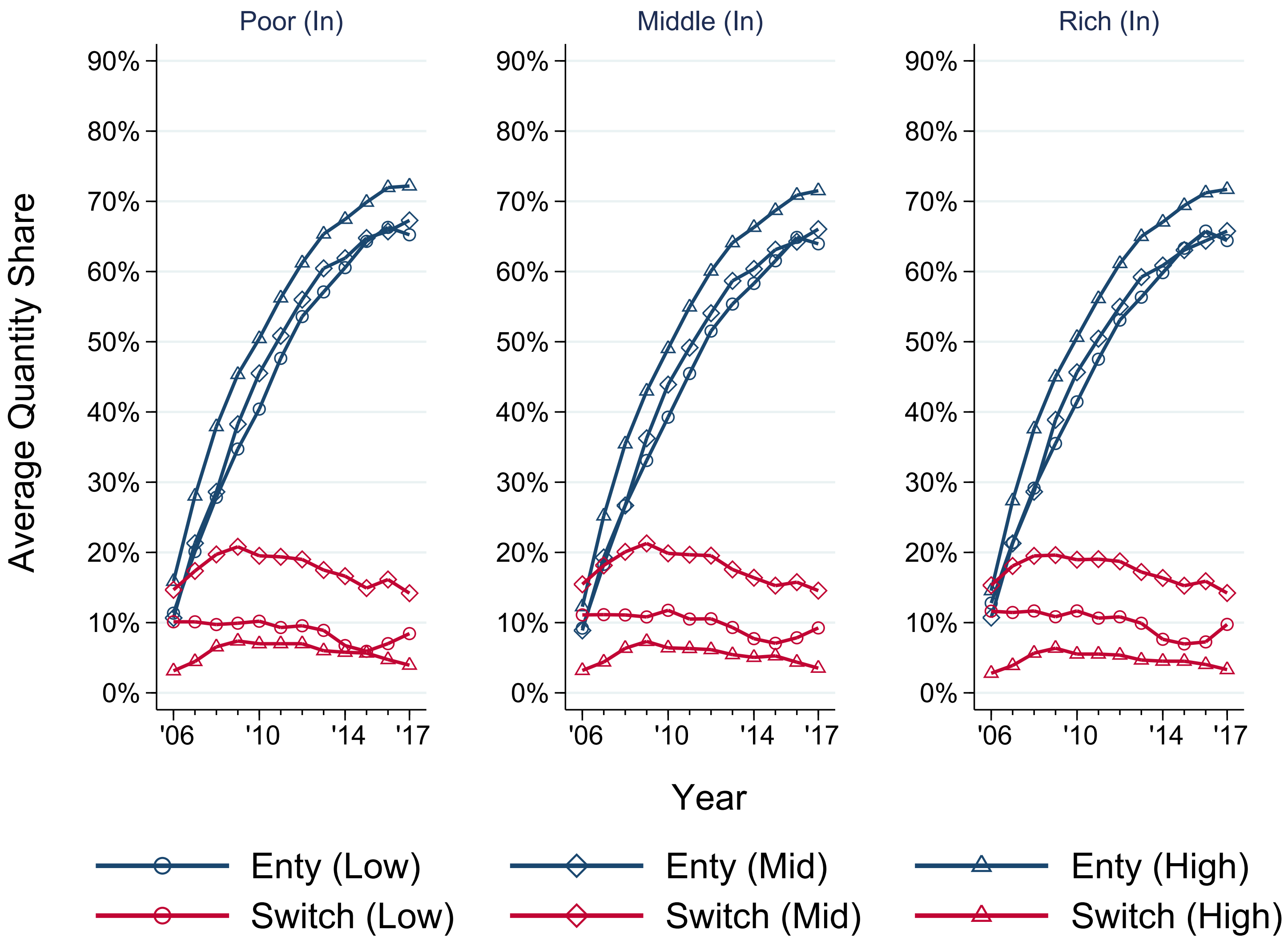
Note: The computation is based on the formulas in equations (7)–(11), relative to base year 2005. The price groups $p \in \{L, M, H\}$ are based on tercile price cutoffs. Data are from Nielsen Consumer Panel (2005–2017).

Figure 11: Proportion of Entry and Switch-in Products/Varieties



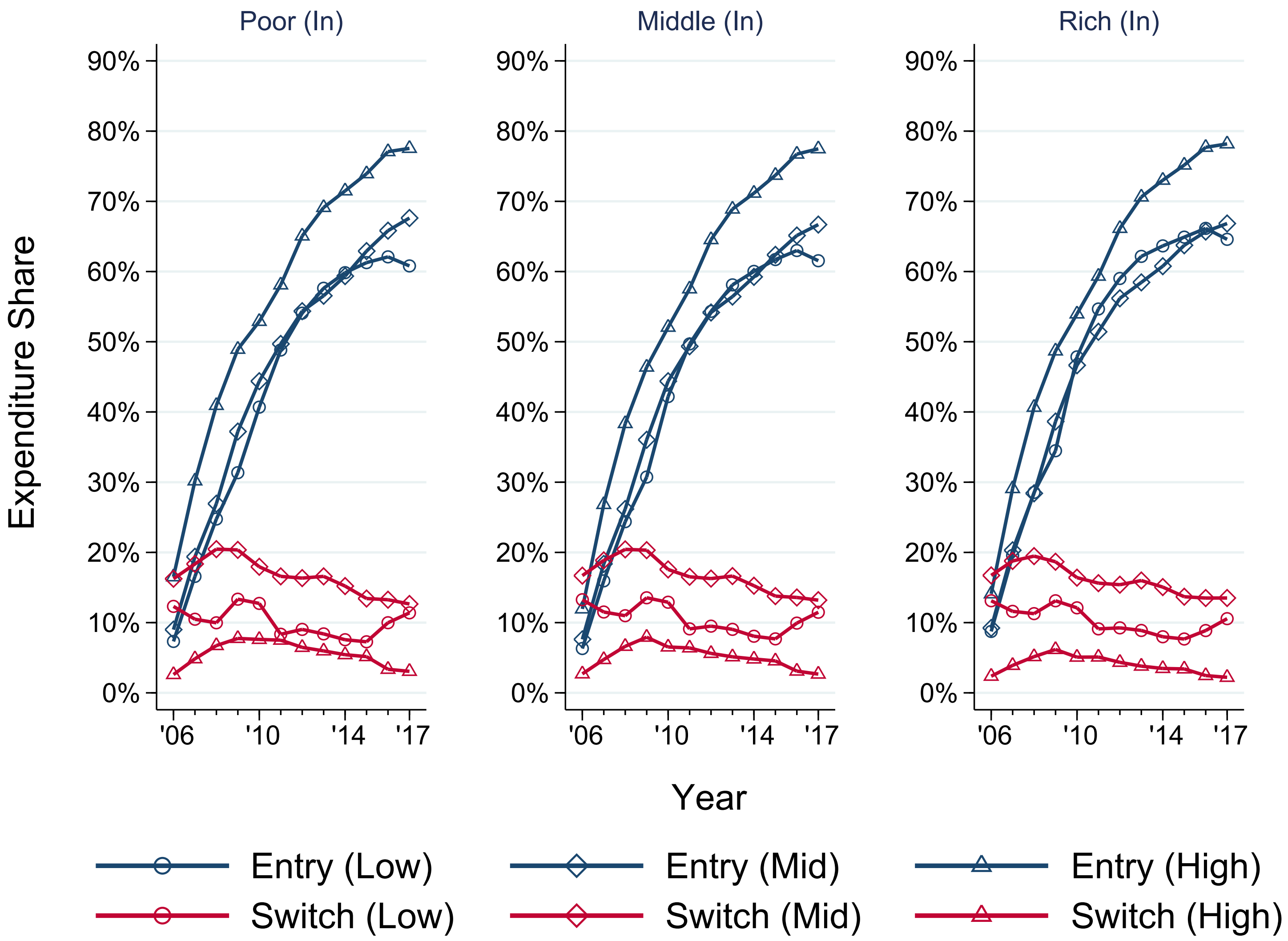
Note: We construct the proportion of entry/switch-in varieties using the number of these varieties divided by the total number of varieties in the current year t for each price-income group, by pulling across all PGCs. The price groups $p \in \{L, M, H\}$ are based on tercile price cutoffs. Data are from Nielsen Consumer Panel (2005–2017).

Figure 12: Quantity Share of Entry and Switch-in Products/Varieties



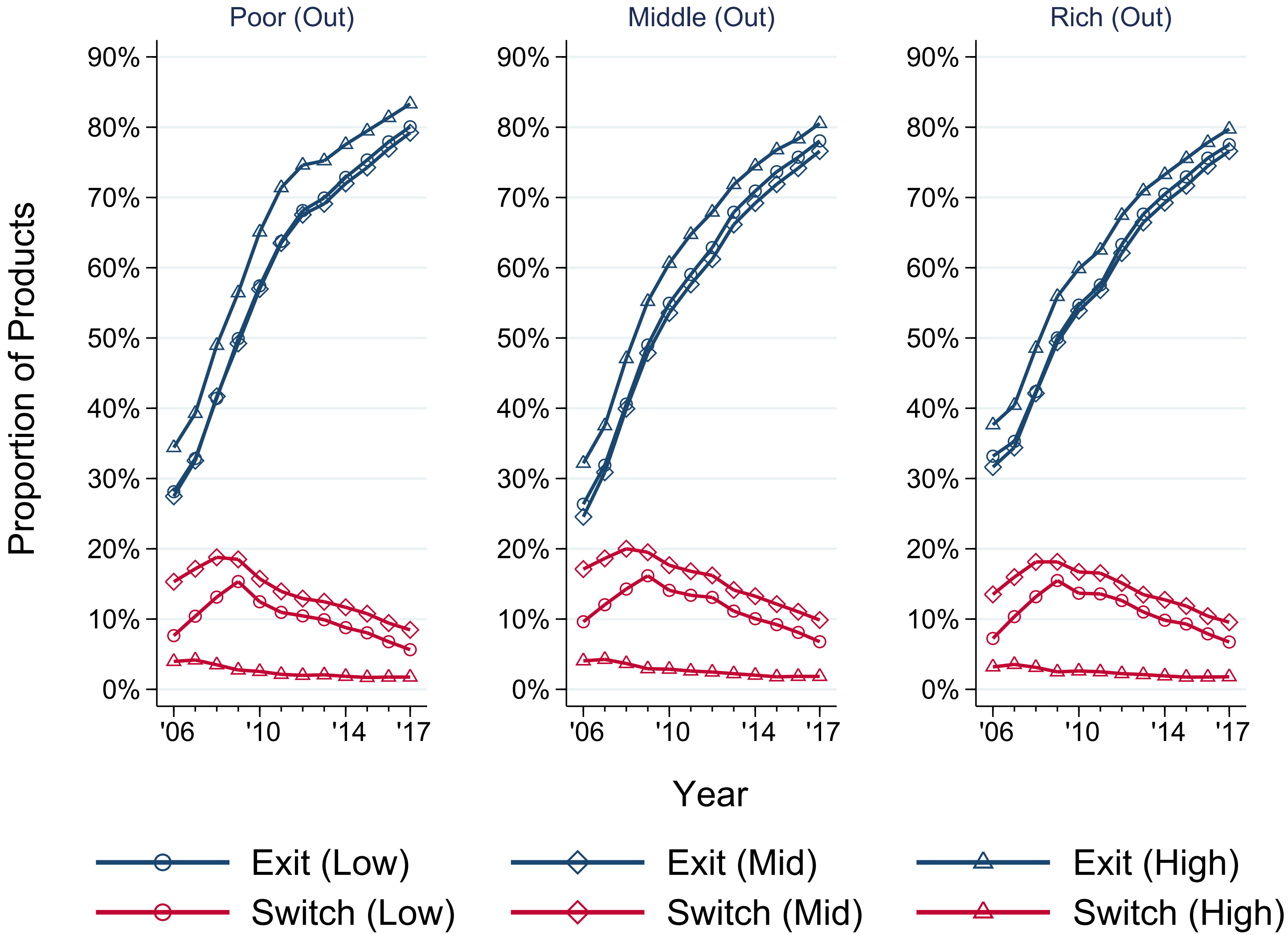
Note: We compute the quantity share of entry/switch-in products for each price-income group and PGC, relative to the current year t, and take the average of the shares across all PGCs. This is in view that the physical units of measure are not the same across PGCs (while all varieties within a PGC have the same physical unit of measure by construction, cf. Section 2). The price groups $p \in \{L, M, H\}$ are based on tercile price cutoffs. Data are from Nielsen Consumer Panel (2005–2017).

Figure 13: Expenditure Share of Entry and Switch-in Products/Varieties



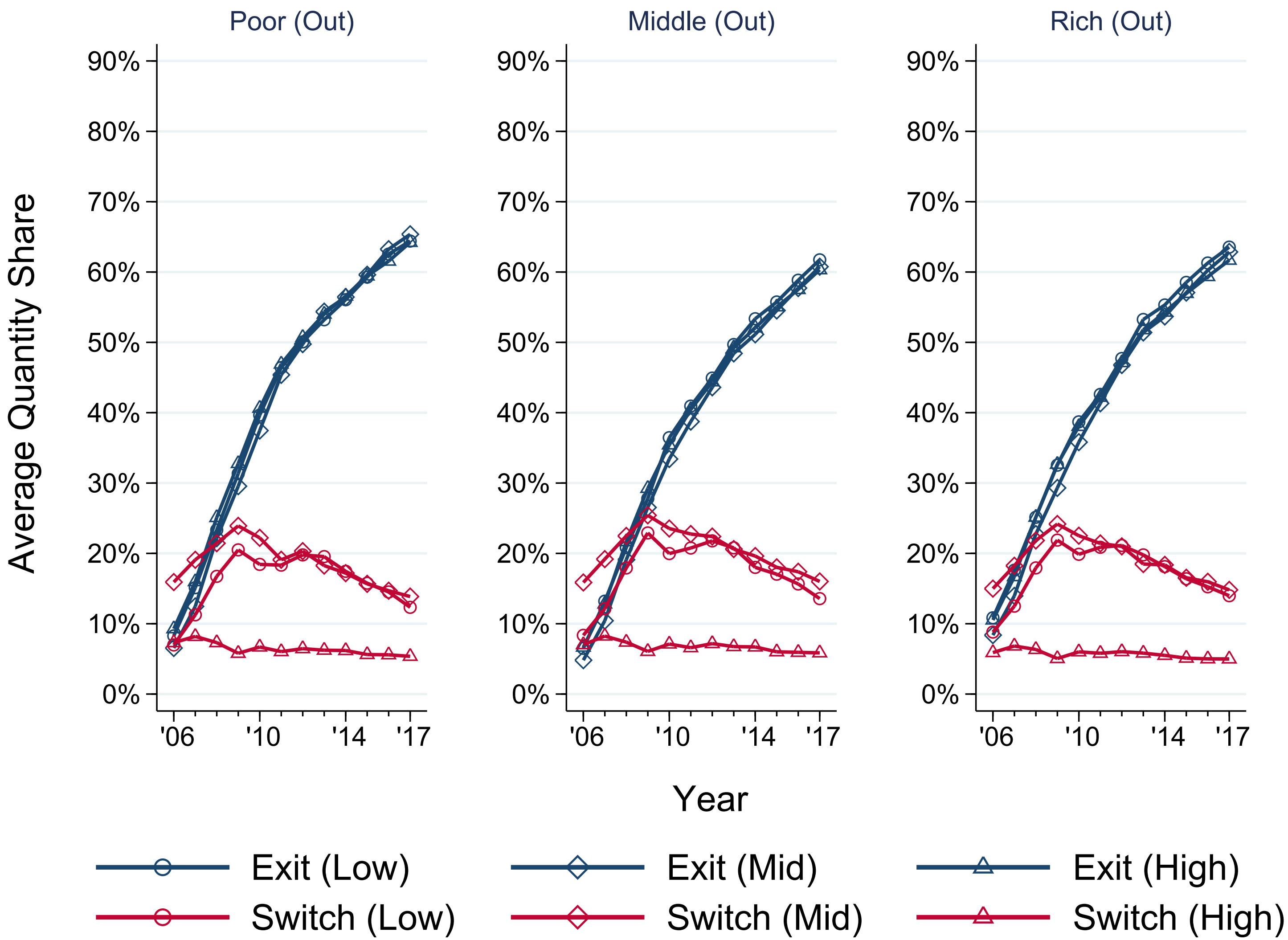
Note: We construct the expenditure share of entry/switch-in varieties for each price-income group, relative to the current year t, by pulling the expenditures across PGCs. The price groups $p \in \{L, M, H\}$ are based on tercile price cutoffs. Data are from Nielsen Consumer Panel (2005–2017).

Figure 14: Proportion of Exit and Switch-out Products/Varieties



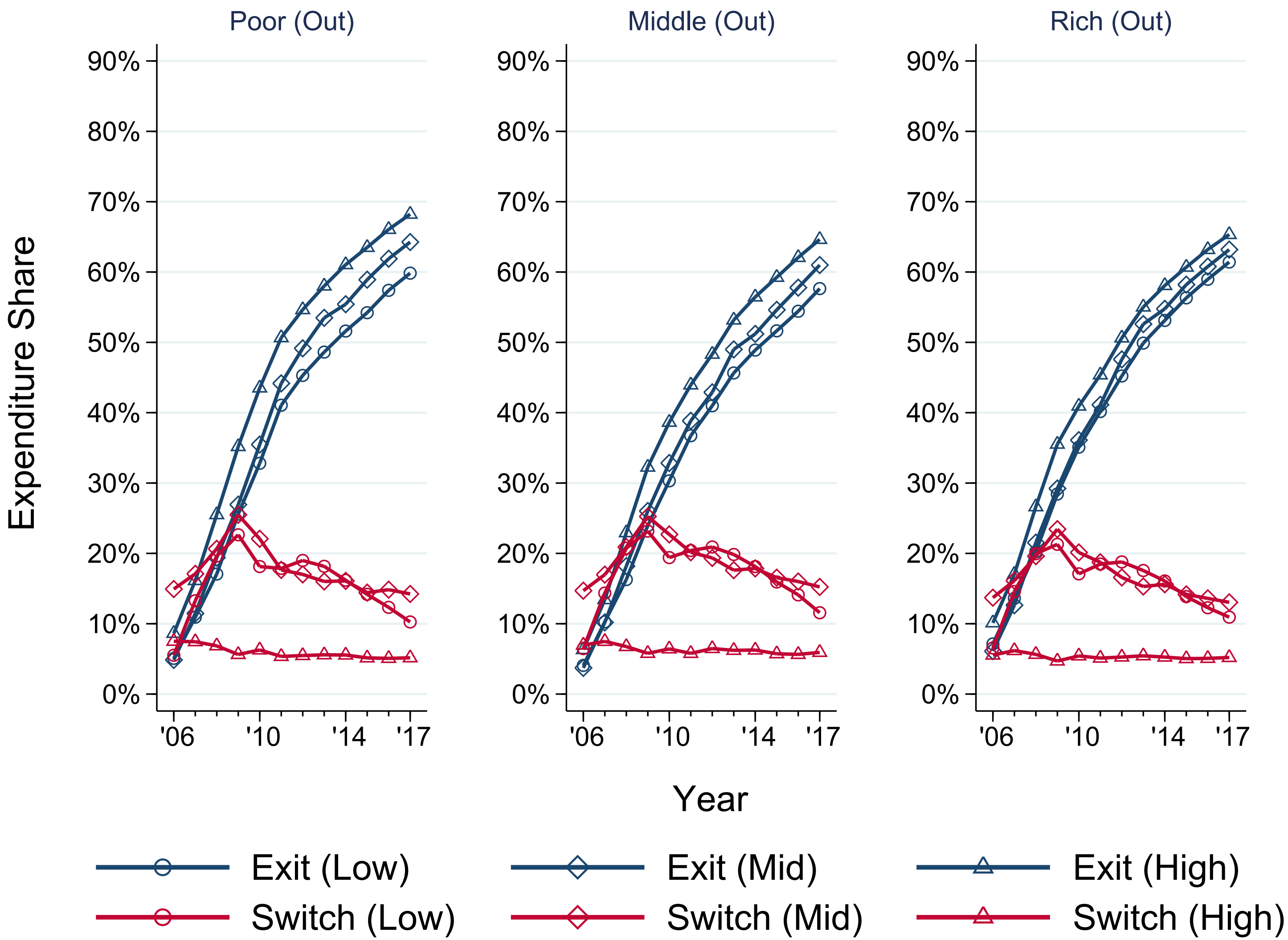
Note: We construct the proportion of exit/switch-out varieties using the number of these varieties divided by the total number of varieties in the base year 2005 for each price-income group, by pulling across all PGCs. The price groups $p \in \{L, M, H\}$ are based on tercile price cutoffs. Data are from Nielsen Consumer Panel (2005–2017).

Figure 15: Quantity Share of Exit and Switch-out Products/Varieties



Note: We compute the quantity shares of exit/switch-out products for each price-income group and PGC, relative to the base year 2005, and take the average of the shares across all PGCs. This is in view that the physical units of measure are not the same across PGCs (while all varieties within a PGC have the same physical unit of measure by construction, cf. Section 2). The price groups $p \in \{L, M, H\}$ are based on tercile price cutoffs. Data are from Nielsen Consumer Panel (2005–2017).

Figure 16: Expenditure Share of Exit and Switch-out Products/Varieties



Note: We construct the expenditure share of exit/switch-out varieties for each price-income group, relative to the base year 2005, by pulling the expenditures across PGCs. The price groups $p \in \{L, M, H\}$ are based on tercile price cutoffs. Data are from Nielsen Consumer Panel (2005–2017).

Figure 17: Bias in Cost-of-Living (with versus without non-homothetic preferences across price groups)

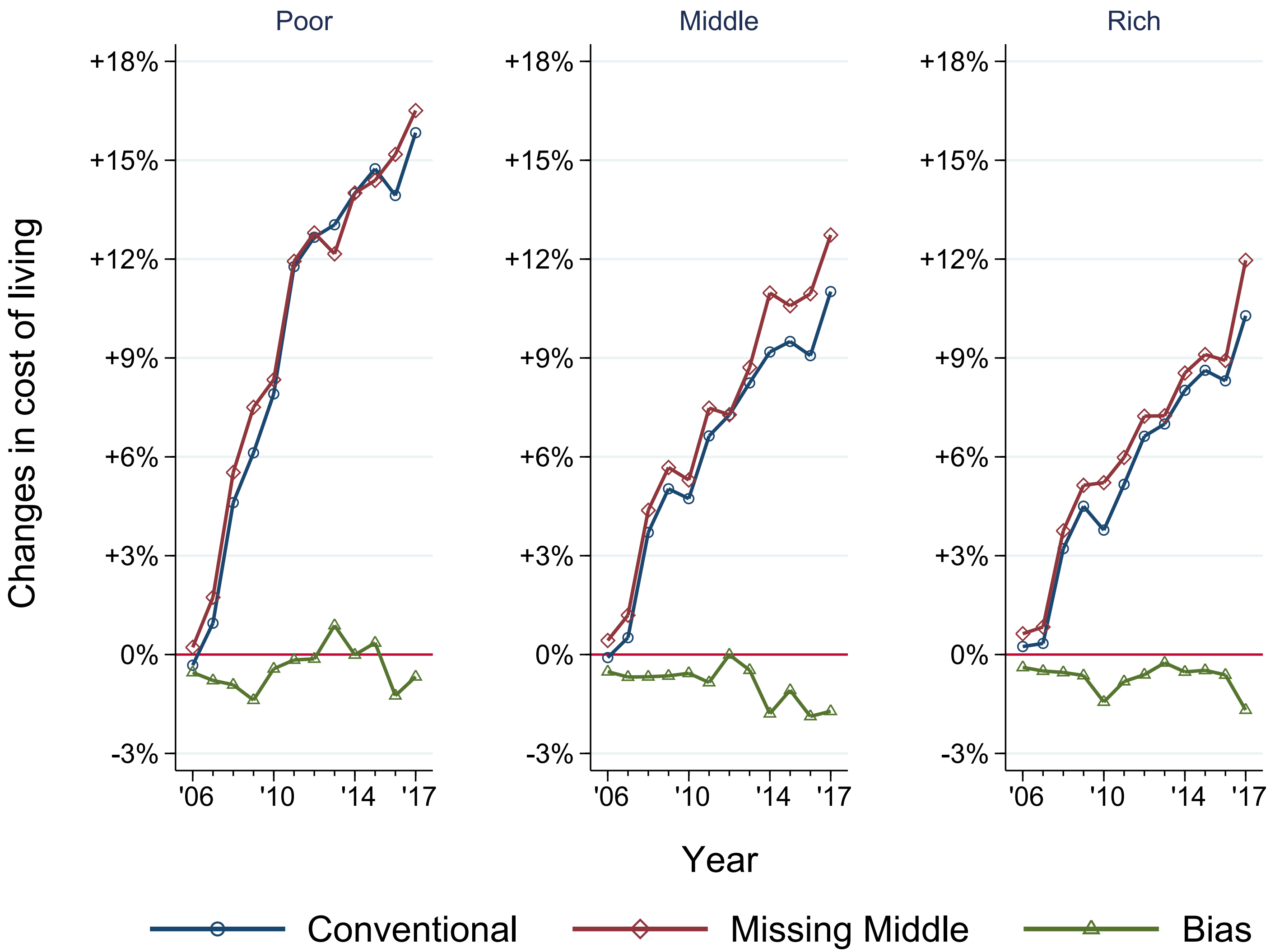


Figure 18: Bias in Entry Effects (with versus without non-homothetic preferences across price groups)

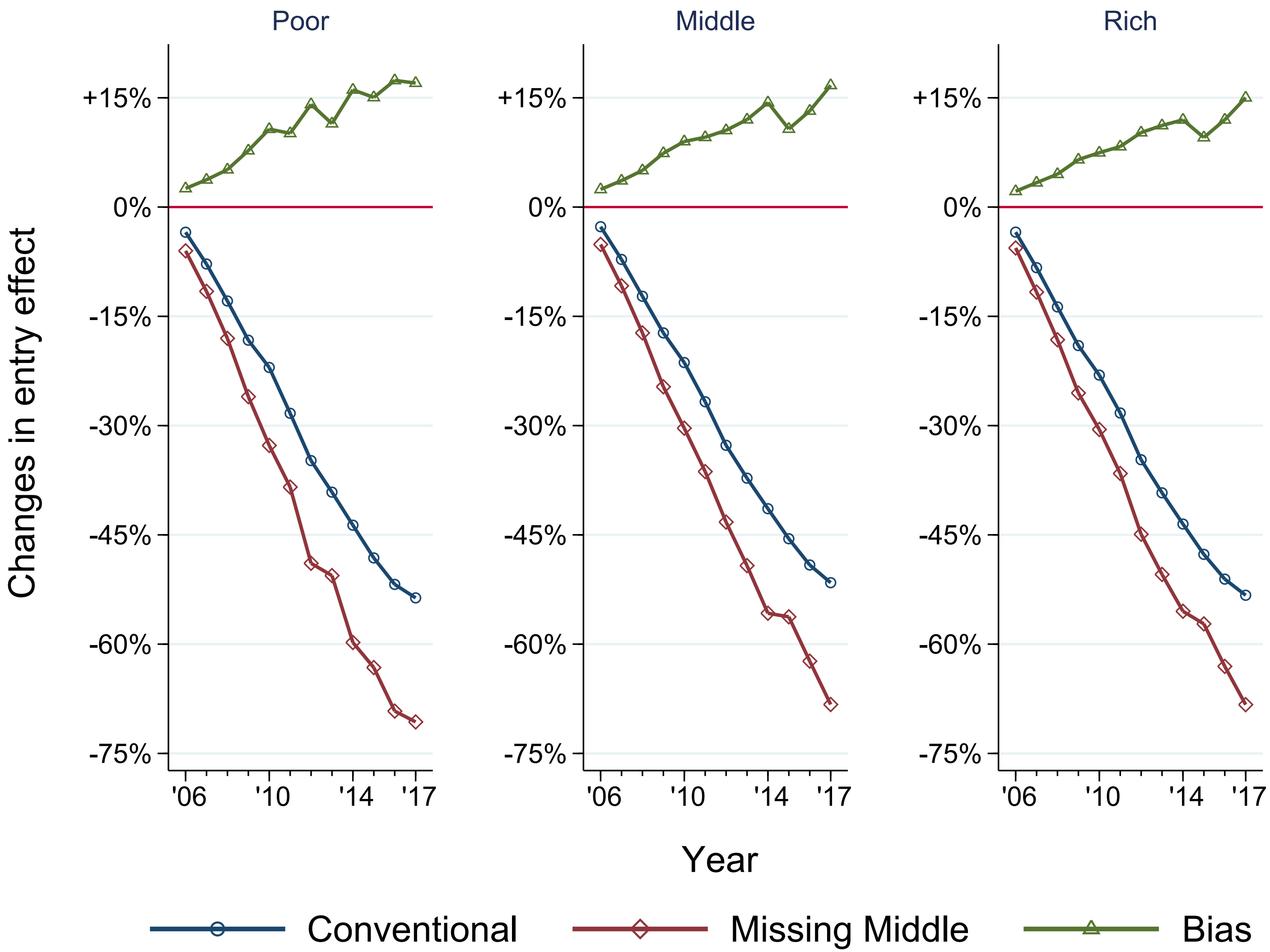


Figure 19: Bias in Exit Effects (with versus without non-homothetic preferences across price groups)

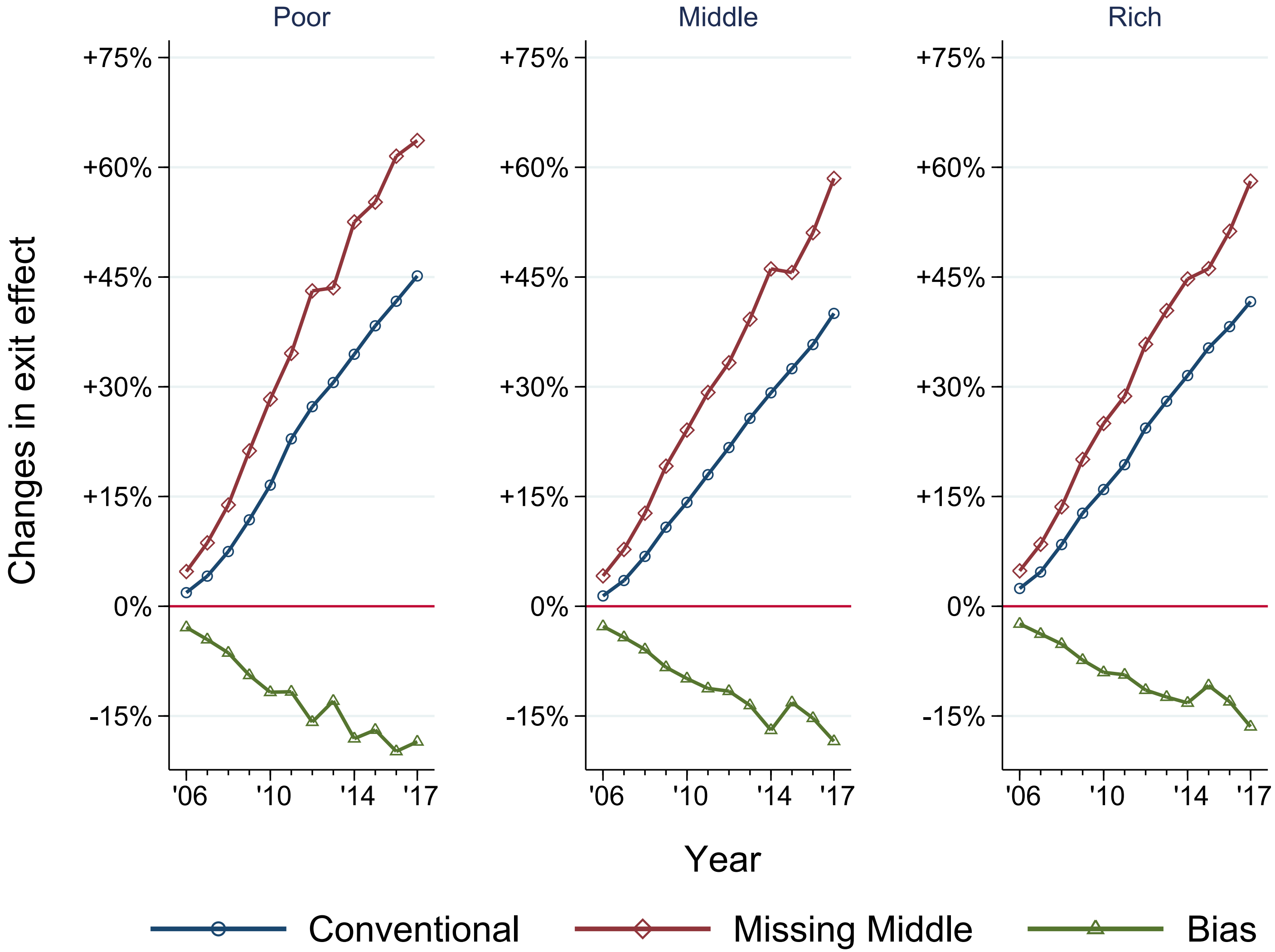


Figure 20: Bias in Net-Entry Effects (with versus without non-homothetic preferences across price groups)

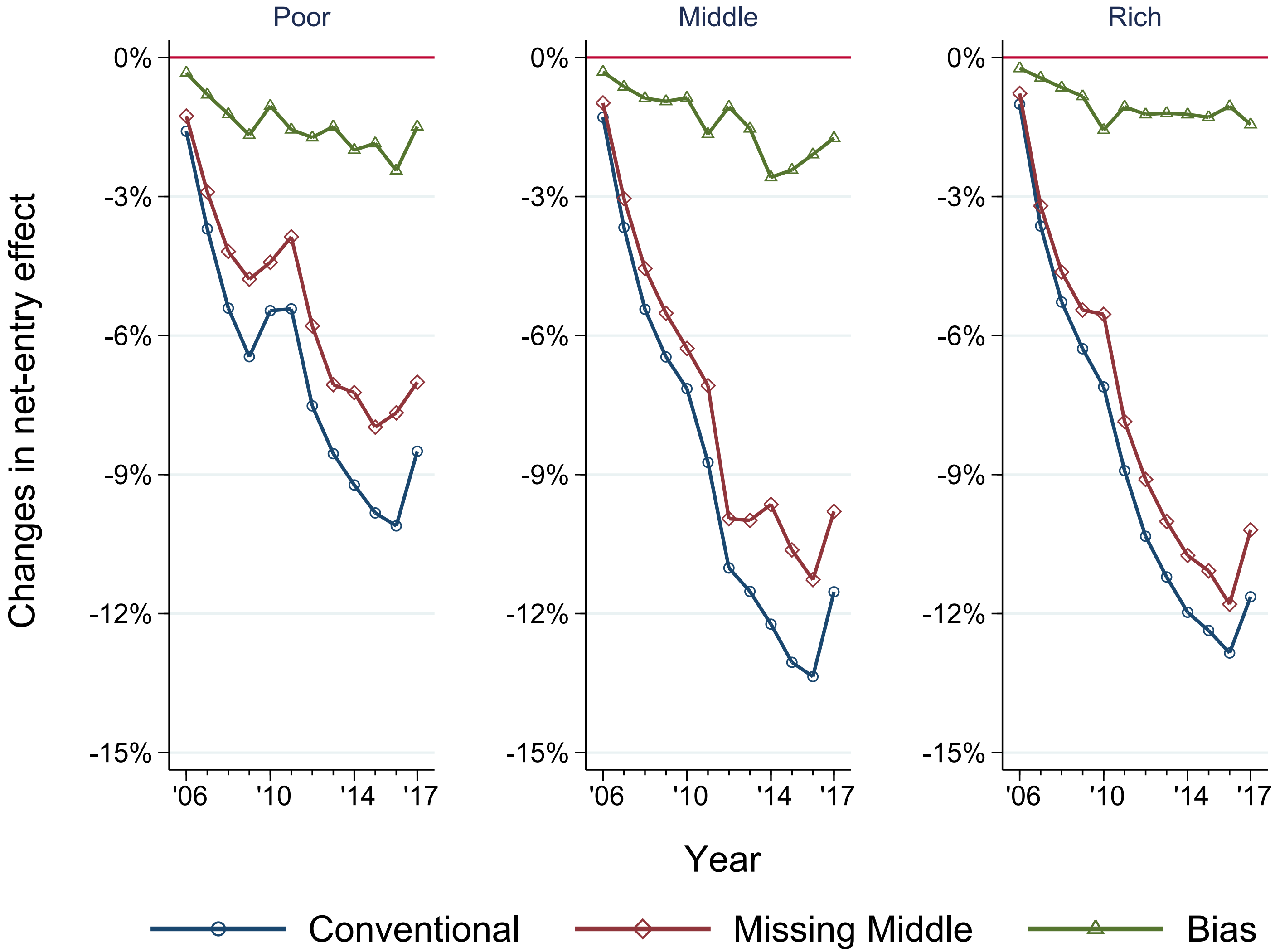


Figure 21: Bias in Pro-Competitive Price Effects (with versus without non-homothetic preferences across price groups)

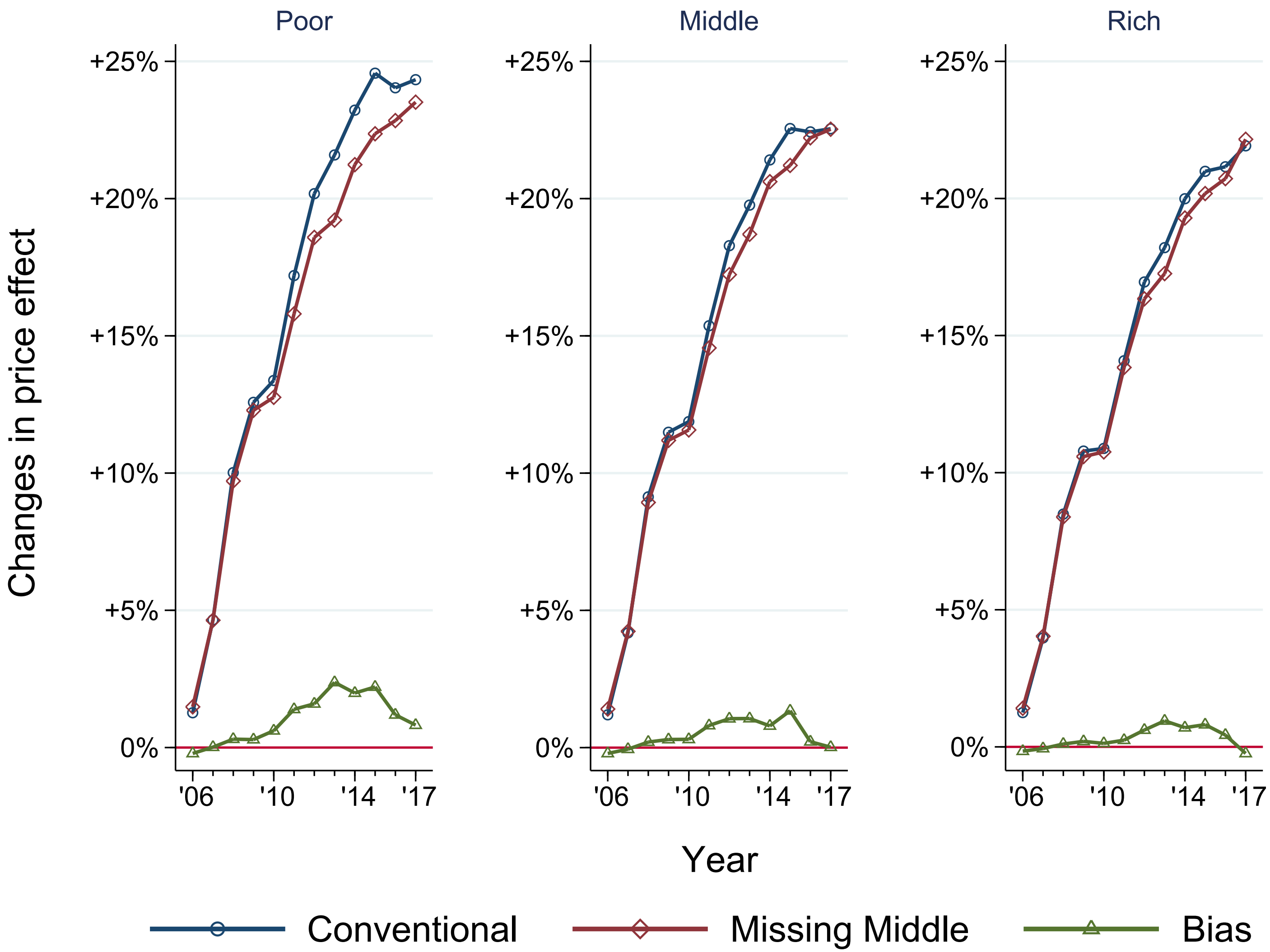


Table A.1: Product Market Shares on Lagged Population Shares — State-Level; Tercile Price Cutoffs; Adjusted Per-Capita Income

	$Share_{s,t,g,L}$ (1)	$Share_{s,t,g,M}$ (2)	$Share_{s,t,g,H}$ (3)	$Share_{s,t,g,L}$ (4)	$Share_{s,t,g,M}$ (5)	$Share_{s,t,g,H}$ (6)
$Frac_{s,t-1,L}$	0.0706			0.067*		
	(0.046)			(0.037)		
$Frac_{s,t-1,M}$		0.326***			0.322***	
		(0.038)			(0.034)	
$Frac_{s,t-1,H}$			0.673***			0.649***
			(0.051)			(0.040)
Robust SE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Product FE	No	No	No	Yes	Yes	Yes
R^2	0.015	0.005	0.014	0.441	0.299	0.514
N	57,218	57,218	57,218	57,218	57,218	57,218

Notes: Market share measures are constructed using the Nielsen Retail Scanner Data (2006–2017) as documented in Section 2.1. Population share measures are constructed using the IPUMS ACS Data (2005–2017) as documented in Section 3.1. States include all the contiguous states in the United States and the District of Columbia. All regressions include a constant term. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. In this exercise, the L-priced (H-priced) products are defined as those UPCs whose deflated real price is below the 33rd (above the 66th) percentile of the year-2006 price distribution (of the corresponding PGC g in area s). The M-priced products are defined as those with real prices in between the two terciles. Income per capita is calculated based on formula (A1).

Table A.2: Product Market Shares on Lagged Population Shares — State-Level; Quartile Price Cutoffs; Adjusted Per-Capita Income

	$Share_{s,t,g,L}$ (1)	$Share_{s,t,g,M}$ (2)	$Share_{s,t,g,H}$ (3)	$Share_{s,t,g,L}$ (4)	$Share_{s,t,g,M}$ (5)	$Share_{s,t,g,H}$ (6)
$Frac_{s,t-1,L}$	0.004			0.002		
	(0.042)			(0.034)		
$Frac_{s,t-1,M}$		0.423***			0.421***	
		(0.041)			(0.035)	
$Frac_{s,t-1,H}$			0.710***			0.686***
			(0.049)			(0.037)
Robust SE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Product FE	No	No	No	Yes	Yes	Yes
R^2	0.014	0.005	0.016	0.430	0.349	0.532
N	57,218	57,218	57,218	57,218	57,218	57,218

Notes: See the footnote of Table A.1. In this exercise, the L-priced (H-priced) products are defined as those UPCs whose deflated real price is below the 25th (above the 75th) percentile of the year-2006 price distribution (of the corresponding PGC g in area s). The M-priced products are defined as those with real prices in between the two quartiles. Income per capita is calculated based on formula (A1).

Table A.3: Product Market Shares on Lagged Population Shares — State-Level; Decile Price Cutoffs; Adjusted Per-Capita Income

	$Share_{s,t,g,L}$ (1)	$Share_{s,t,g,M}$ (2)	$Share_{s,t,g,H}$ (3)	$Share_{s,t,g,L}$ (4)	$Share_{s,t,g,M}$ (5)	$Share_{s,t,g,H}$ (6)
$Frac_{s,t-1,L}$	0.033			0.032		
	(0.030)			(0.026)		
$Frac_{s,t-1,M}$		0.374***			0.374***	
		(0.034)			(0.029)	
$Frac_{s,t-1,H}$			0.528***			0.512***
			(0.033)			(0.025)
Robust SE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Product FE	No	No	No	Yes	Yes	Yes
R^2	0.008	0.005	0.015	0.366	0.376	0.539
N	57,218	57,218	57,218	57,218	57,218	57,218

Notes: See the footnote of Table A.1. In this exercise, the L-priced (H-priced) products are defined as those UPCs whose deflated real price is below the 10th (above the 90th) percentile of the year-2006 price distribution (of the corresponding PGC g in area s). The M-priced products are defined as those with real prices in between the two deciles. Income per capita is calculated based on formula (A1).

Table A.4: Product Market Shares on Lagged Population Shares — Commuting-Zone-Level; Tercile Price Cutoffs; Adjusted Per-Capita Income

	$Share_{s,t,g,L}$ (1)	$Share_{s,t,g,M}$ (2)	$Share_{s,t,g,H}$ (3)	$Share_{s,t,g,L}$ (4)	$Share_{s,t,g,M}$ (5)	$Share_{s,t,g,H}$ (6)
$Frac_{s,t-1,L}$	0.127***			0.117***		
	(0.010)			(0.008)		
$Frac_{s,t-1,M}$		0.100***			0.104***	
		(0.009)			(0.008)	
$Frac_{s,t-1,H}$			0.578***			0.580***
			(0.016)			(0.013)
Robust SE	Yes	Yes	Yes	Yes	Yes	Yes
CZone FE	Yes	Yes	Yes	Yes	Yes	Yes
Product FE	No	No	No	Yes	Yes	Yes
R^2	0.234	0.077	0.103	0.491	0.275	0.466
N	642,031	642,031	642,031	642,031	642,031	642,031

Notes: Market share measures are constructed using the Nielsen Retail Scanner Data (2006–2017) as documented in Section 2.1. Population share measures are constructed using the IPUMS ACS Data (2005–2017) as documented in Section 3.1. Note that the Nielsen Retail Scanner Data do not report sales data for all 722 commuting zones in the contiguous U.S., especially in those least populated areas. In particular, the Nielsen Retail Scanner Data cover 660 commuting zones in 2006 (min), 690 in 2016 (max), and 688 in 2017. All of the missing commuting zones belong to Bin 1 in Table 2, except one that belongs to Bin 2. All regressions include a constant term. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. In this exercise, the L-priced (H-priced) products are defined as those UPCs whose deflated real price is below the 33rd (above the 66th) percentile of the year-2006 price distribution (of the corresponding PGC g in area s). The M-priced products are defined as those with real prices in between the two terciles. Income per capita is calculated based on formula (A1).

Table A.5: Product Market Shares on Lagged Population Shares — Commuting-Zone-Level; Quartile Price Cutoffs; Adjusted Per-Capita Income

	$Share_{s,t,g,L}$ (1)	$Share_{s,t,g,M}$ (2)	$Share_{s,t,g,H}$ (3)	$Share_{s,t,g,L}$ (4)	$Share_{s,t,g,M}$ (5)	$Share_{s,t,g,H}$ (6)
$Frac_{s,t-1,L}$	0.112***			0.103***		
	(0.009)			(0.008)		
$Frac_{s,t-1,M}$		0.135***			0.143***	
		(0.010)			(0.009)	
$Frac_{s,t-1,H}$			0.603***			0.608***
			(0.015)			(0.012)
Robust SE	Yes	Yes	Yes	Yes	Yes	Yes
CZone FE	Yes	Yes	Yes	Yes	Yes	Yes
Product FE	No	No	No	Yes	Yes	Yes
R^2	0.290	0.131	0.083	0.513	0.365	0.473
N	642,031	642,031	642,031	642,031	642,031	642,031

Notes: See the footnote of Table A.4. In this exercise, the L-priced (H-priced) products are defined as those UPCs whose deflated real price is below the 25th (above the 75th) percentile of the year-2006 price distribution (of the corresponding PGC g in area s). The M-priced products are defined as those with real prices in between the two quartiles. Income per capita is calculated based on formula (A1).

Table A.6: Product Market Shares on Lagged Population Shares — Commuting-Zone-Level; Decile Price Cutoffs; Adjusted Per-Capita Income

	$Share_{s,t,g,L}$ (1)	$Share_{s,t,g,M}$ (2)	$Share_{s,t,g,H}$ (3)	$Share_{s,t,g,L}$ (4)	$Share_{s,t,g,M}$ (5)	$Share_{s,t,g,H}$ (6)
$Frac_{s,t-1,L}$	0.056***			0.050***		
	(0.008)			(0.008)		
$Frac_{s,t-1,M}$		0.129***			0.142***	
		(0.010)			(0.008)	
$Frac_{s,t-1,H}$			0.451***			0.465***
			(0.011)			(0.009)
Robust SE	Yes	Yes	Yes	Yes	Yes	Yes
CZone FE	Yes	Yes	Yes	Yes	Yes	Yes
Product FE	No	No	No	Yes	Yes	Yes
R^2	0.453	0.296	0.058	0.577	0.530	0.521
N	642,031	642,031	642,031	642,031	642,031	642,031

Notes: See the footnote of Table A.4. In this exercise, the L-priced (H-priced) products are defined as those UPCs whose deflated real price is below the 10th (above the 90th) percentile of the year-2006 price distribution (of the corresponding PGC g in area s). The M-priced products are defined as those with real prices in between the two deciles. Income per capita is calculated based on formula (A1).

Table A.7: Product Market Shares on Lagged Population Shares ($\times\mathbf{1}\{$Population Density $\geq$ National Median$\}$) — Commuting-Zone-Level; Tercile Price Cutoffs; Adjusted Per-Capita Income

	$Share_{s,t,g,L}$ (1)	$Share_{s,t,g,M}$ (2)	$Share_{s,t,g,H}$ (3)	$Share_{s,t,g,L}$ (4)	$Share_{s,t,g,M}$ (5)	$Share_{s,t,g,H}$ (6)
$Frac_{s,t-1,L}$	0.140***			0.129***		
	(0.010)			(0.009)		
$Frac_{s,t-1,L}$	-0.373***			-0.358***		
$\times\mathbf{1}\{Density_s \geq median\}$	(0.049)			(0.039)		
$Frac_{s,t-1,M}$		0.077***			0.082***	
		(0.010)			(0.009)	
$Frac_{s,t-1,M}$		0.319***			0.313***	
$\times\mathbf{1}\{Density_s \geq median\}$		(0.033)			(0.029)	
$Frac_{s,t-1,H}$			0.564***			0.566***
			(0.016)			(0.013)
$Frac_{s,t-1,H}$			0.247***			0.238***
$\times\mathbf{1}\{Density_s \geq median\}$			(0.063)			(0.047)
Robust SE	Yes	Yes	Yes	Yes	Yes	Yes
CZone FE	Yes	Yes	Yes	Yes	Yes	Yes
Product FE	No	No	No	Yes	Yes	Yes
R^2	0.234	0.077	0.103	0.491	0.275	0.466
N	642,031	642,031	642,031	642,031	642,031	642,031

Notes: See the footnote of Table A.4. We compute the population density for each CZ, using a crosswalk from the county level to the CZ level provided by Autor and Dorn (2013), the county population in 2005 provided by the NBER, and the county land areas provided by the Census Bureau. The indicator $\mathbf{1}\{Density_s \geq median\}$ equals one if the population density of CZ s is above or equal to the national median (across all 722 commuting zones). In this exercise, the L-priced (H-priced) products are defined as those UPCs whose deflated real price is below the 33rd (above the 66th) percentile of the year-2006 price distribution (of the corresponding PGC g in area s). The M-priced products are defined as those with real prices in between the two terciles. Income per capita is calculated based on formula (A1).

Table A.8: Product Market Shares on Lagged Population Shares ($\times\mathbf{1}\{$Population Density $\geq$ National Median$\}$) — Commuting-Zone-Level; Quartile Price Cutoffs; Adjusted Per-Capita Income

	$Share_{s,t,g,L}$ (1)	$Share_{s,t,g,M}$ (2)	$Share_{s,t,g,H}$ (3)	$Share_{s,t,g,L}$ (4)	$Share_{s,t,g,M}$ (5)	$Share_{s,t,g,H}$ (6)
$Frac_{s,t-1,L}$	0.122***			0.112***		
	(0.009)			(0.008)		
$Frac_{s,t-1,L}$	-0.279***			-0.266***		
$\times\mathbf{1}\{Density_s \geq median\}$	(0.046)			(0.037)		
$Frac_{s,t-1,M}$		0.105***			0.113***	
		(0.011)			(0.009)	
$Frac_{s,t-1,M}$		0.432***			0.423***	
$\times\mathbf{1}\{Density_s \geq median\}$		(0.036)			(0.031)	
$Frac_{s,t-1,H}$			0.586***			0.592***
			(0.015)			(0.012)
$Frac_{s,t-1,H}$			0.284***			0.272***
$\times\mathbf{1}\{Density_s \geq median\}$			(0.060)			(0.044)
Robust SE	Yes	Yes	Yes	Yes	Yes	Yes
CZone FE	Yes	Yes	Yes	Yes	Yes	Yes
Product FE	No	No	No	Yes	Yes	Yes
R^2	0.290	0.132	0.083	0.513	0.365	0.473
N	642,031	642,031	642,031	642,031	642,031	642,031

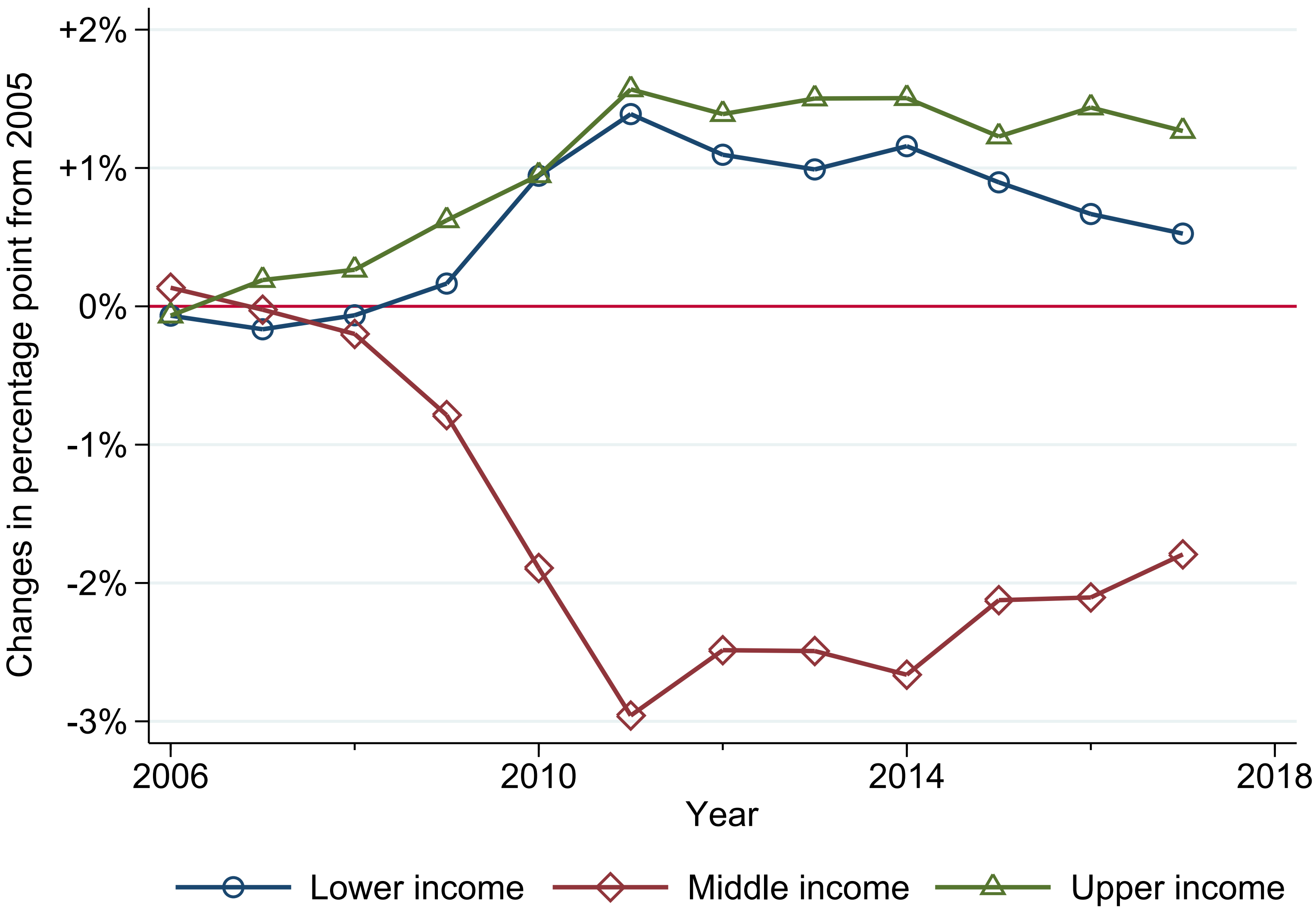
Notes: See the footnote of Table A.7. In this exercise, the L-priced (H-priced) products are defined as those UPCs whose deflated real price is below the 25th (above the 75th) percentile of the year-2006 price distribution (of the corresponding PGC g in area s). The M-priced products are defined as those with real prices in between the two quartiles. Income per capita is calculated based on formula (A1).

Table A.9: Product Market Shares on Lagged Population Shares ($\times\mathbf{1}\{$Population Density $\geq$ National Median$\}$) — Commuting-Zone-Level; Decile Price Cutoffs; Adjusted Per-Capita Income

	$Share_{s,t,g,L}$ (1)	$Share_{s,t,g,M}$ (2)	$Share_{s,t,g,H}$ (3)	$Share_{s,t,g,L}$ (4)	$Share_{s,t,g,M}$ (5)	$Share_{s,t,g,H}$ (6)
$Frac_{s,t-1,L}$	0.056***			0.049***		
	(0.008)			(0.007)		
$Frac_{s,t-1,L}$	0.014***			0.018***		
$\times\mathbf{1}\{Density_s \geq median\}$	(0.036)			(0.031)		
$Frac_{s,t-1,M}$		0.099***			0.113***	
		(0.010)			(0.009)	
$Frac_{s,t-1,M}$		0.411***			0.408***	
$\times\mathbf{1}\{Density_s \geq median\}$		(0.036)			(0.027)	
$Frac_{s,t-1,H}$			0.443***			0.458***
			(0.012)			(0.009)
$Frac_{s,t-1,H}$			0.123***			0.130***
$\times\mathbf{1}\{Density_s \geq median\}$			(0.050)			(0.029)
Robust SE	Yes	Yes	Yes	Yes	Yes	Yes
CZone FE	Yes	Yes	Yes	Yes	Yes	Yes
Product FE	No	No	No	Yes	Yes	Yes
R^2	0.453	0.296	0.058	0.577	0.530	0.521
N	642,031	642,031	642,031	642,031	642,031	642,031

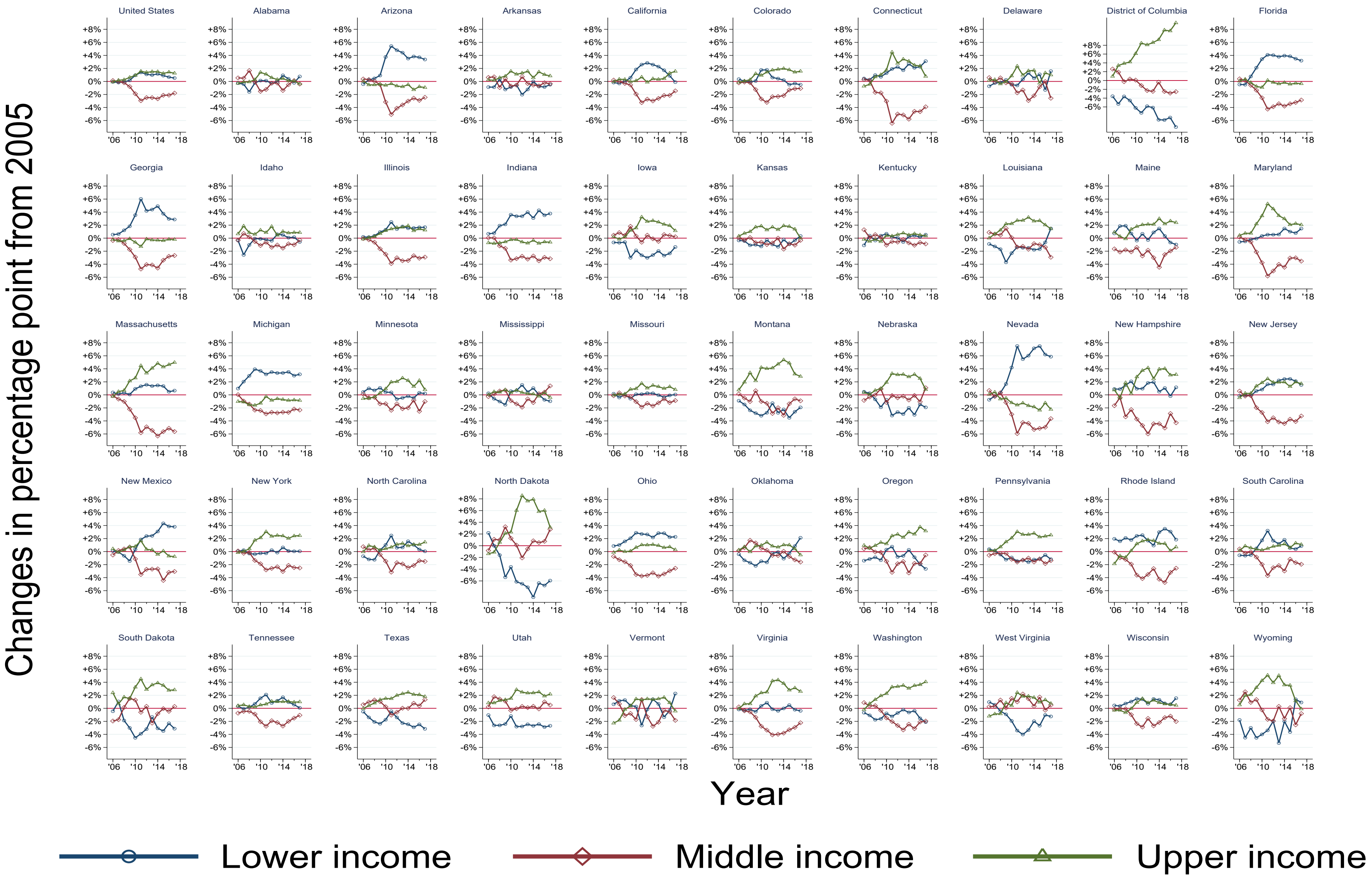
Notes: See the footnote of Table A.7. In this exercise, the L-priced (H-priced) products are defined as those UPCs whose deflated real price is below the 10th (above the 90th) percentile of the year-2006 price distribution (of the corresponding PGC g in area s). The M-priced products are defined as those with real prices in between the two deciles. Income per capita is calculated based on formula (A1).

Figure A.1: Changes in Population Shares of Different Income Classes in the United States (relative to 2005) — Adjusted Per-Capita Income



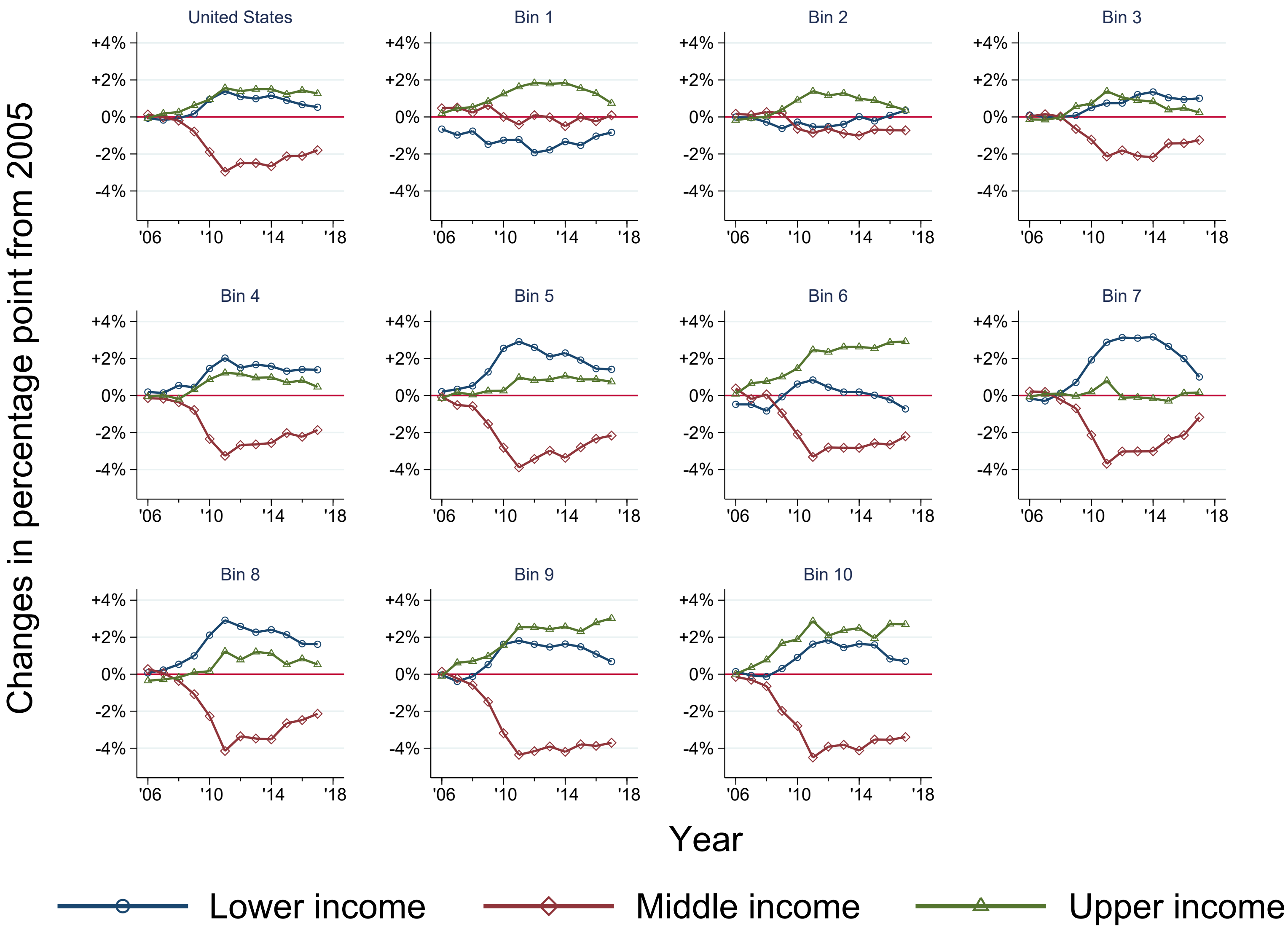
Note: Each plotted point represents the change in the population share (in percentage points) of a particular income group in a year, relative to 2005, for the United States. Income per capita is calculated based on formula (A1). Data are from the IPUMS ACS (2005–2017).

Figure A.2: Changes in Population Shares of Different Income Classes by States/District — Adjusted Per-Capita Income



Note: Each plotted point represents the change in the population share (in percentage points) of a particular income group in a year, relative to 2005, for a state/district. Income per capita is calculated based on formula (A1). Data are from the IPUMS ACS (2005–2017).

Figure A.3: Changes in Population Shares of Different Income Classes by Population Density of Commuting Zones — Adjusted Per-Capita Income



Note: Each plotted point represents the change in the population share (in percentage points) of a particular income group in a year, relative to 2005, for the commuting zones within the same bin of population density. Bin 10 (1) has the highest (lowest) population density. Income per capita is calculated based on formula (A1). Data are from the IPUMS ACS (2005–2017).